Using Empirical Data to Evaluate Strategies to Improve Women's Health

Using empirical data to evaluate strategies to improve women's health

A dissertation presented

by

Arielle Gwendolyn Bensimon

to

The Committee on Higher Degrees in Health Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Health Policy

Harvard University

Cambridge, Massachusetts

May 2016

Dissertation Advisor:
Professor Jane Kim

Author:
Arielle Gwendolyn Bensimon

Using empirical data to evaluate strategies to improve women's health

Abstract

My three papers evaluate the effectiveness and cost-effectiveness of clinical and policy strategies to improve women's health, focusing on human papillomavirus (HPV) vaccination in the U.S. and maternal health care in a developing country context.

Paper 1 presents a claims-based econometric analysis of the U.S.'s Patient Protection and Affordable Care Act provision requiring the elimination of cost-sharing for recommended preventive care. I evaluate the effect of this value-based insurance design intervention on HPV immunization rates among girls and young women enrolled in private insurance plans. My regression approach uses variation in the intensity and timing of the intervention across plans to distinguish policy effects from background trends. I find that the policy was associated with modest increases in age-specific vaccination rates. Increases in vaccination per dollar reduction in cost-sharing were notably larger among beneficiaries in socioeconomically disadvantaged areas. Nevertheless, vaccination rates under free preventive care were well below federal targets, highlighting the need for additional interventions to increase HPV vaccine coverage.

In Paper 2, I undertake a comparative effectiveness analysis of HPV vaccination by dose level within a U.S. cohort of adolescent girls and young women. Rates of screening-detected cervical abnormalities in claims are compared among recipients of zero, one, two, or three doses, using a marginal structural model approach to adjust for a broader set of potential confounders

than would be possible with conventional regression methods. Findings from these analyses complement prior evidence from immunogenicity trials, and although protective effects appear greatest with three doses, support the value of HPV vaccination even when incomplete. Vaccine effect estimates are largest with respect to high-grade lesions that are precursors to cervical cancer.

Using primary data from a randomized experiment, Paper 3 examines the cost-effectiveness of pay-for-performance interventions among obstetric care providers in rural Karnataka, India. I construct a decision analytic model to quantify incremental costs and life years under alternative policy scenarios, combining obstetric complication outcomes and program expenditures from the trial with published evidence on complication-related mortality and medical costs. Results suggest that an incentive program based on input quality is not cost-effective in its current form, but could become economically attractive if program activities can be adjusted to reduce costs while maintaining similar health effects. Performance data collection costs were substantial in this resource-limited setting and represent a key barrier to cost-effectiveness.

Table of Contents

Acknowledgements

I am especially grateful to my advisors, Jane Kim, John Hsu, and Miguel Hernán, for their guidance, mentorship, and thoughtful feedback throughout the research process. I would also like to thank Michael Chernew, Vicki Fung, Teresa Gibson, Mary Elizabeth Landrum, Joe Newhouse, Maggie Price, and Mark Schuster for helpful discussions and suggestions on the claims-based human papillomavirus immunization studies. I am similarly thankful to Manoj Mohanan, Katherine Donato, Margaret Pendzich, Manveen Kohli, and the entire COHESIVE-India team for sharing data and financial records from the Karnataka experiment and providing other invaluable assistance with the provider incentives cost-effectiveness analysis. Additional thanks to Debbie Whitney, Jessica Livingston, and Ayres Heller for being amazing resources during my time in the Health Policy PhD program.

This dissertation research was supported by the National Institute of Child Health and Human Development (R01 HD075121), the Agency for Healthcare Research and Quality (T32HS000055), and Harvard University. Funding for the randomized experiment in Karnataka, on which the cost-effectiveness analysis in Paper 3 is based, was provided by the World Bank, the International Initiative for Impact Evaluation (3ie), the UK Department for International Development, and the State Government of Karnataka. The content is solely my responsibility and does not necessarily represent the official views of these funding organizations.

Lastly, this dissertation could not have been completed without the love, support, and encouragement of my husband, Scott Stransky, and all of my family. This of course includes my lifelong pet turtle, Swimmy, who kept me company on many a late night while writing these papers.

Paper 1:

Effects of a large-scale value-based insurance design intervention on HPV immunization rates: Evidence from U.S. health reform

INTRODUCTION

As of September 23, 2010, the Affordable Care Act (ACA) requires all non-grandfathered commercial health insurance plans to fully cover a range of highly recommended preventive services with zero cost-sharing to patients. Among other types of preventive care, the reform extends to annual wellness visits and routine vaccinations for children and adults within the affected health plans. My paper evaluates the effectiveness of this large-scale value-based insurance design intervention targeting preventive care use, focusing specifically on rates of human papillomavirus (HPV) immunization among age-eligible girls and young women (9 to 26 years old) enrolled in private insurance plans. I also examine sources of heterogeneity in individuals' responsiveness to the cost-sharing reductions, which will suggest the policy's likely effect on socioeconomic disparities in vaccine uptake.

HPV vaccines and barriers to uptake

HPV is the most common sexually transmitted disease in the United States (Weinstock 2004), and is responsible for approximately 18,000 new cancer cases per year among U.S. women (CDC 2012). The quadrivalent and bivalent HPV vaccines prevent infection with the two types (HPV-16 and HPV-18) that cause the majority of cervical and other HPV-associated cancers (CDC 2012). In trials, the vaccines have shown high efficacy in preventing 16/18-related precancerous lesions when administered prior to infection (Lehtinen 2012; Kjaer 2009). Consequently, since 2007, the Advisory Committee on Immunization Practices (ACIP) has recommended routine HPV vaccination for girls and women aged 9 to 26 years, with a target age of 11 to 12 years (Markowitz 2014). Decision models have consistently found routine HPV

vaccination programs to be cost-effective among target-age girls, especially if high coverage can be achieved (Kim & Goldie 2009; Elbasha 2007; Chesson 2008).

However, HPV vaccine coverage in the U.S. remains low relative to other vaccines recommended for adolescents (Reagan-Steiner 2015). In 2013, only 54.6% and 36.4% of 13 to 17 year-old girls had initiated and completed the three-dose HPV vaccine series, respectively (Stokley 2014). Individual-level barriers to HPV vaccination reportedly include misconceptions about HPV risk and the value of vaccination (Holman 2013; Williams 2013; Jain 2009). The perception that HPV vaccines are only needed for sexually active individuals may lead parents and vaccine-eligible young women to refuse or delay vaccine initiation (Stokley 2014; Williams 2013). Concerns about cost are another commonly reported reason for vaccine refusal in surveys of parents, patients, and health care providers (Laz 2013; McCave 2010; Anhang 2011). Completing the recommended three-dose series requires multiple interactions with the health care system and (depending on insurance generosity) out-of-pocket costs of up to $400 plus administration fees. In one survey study, 30% of unvaccinated young women replied that they would accept HPV vaccination if it were provided for free or at low cost, but would not be willing to pay the full price out-of-pocket (Anhang 2011).

At the state level, school immunization mandates for HPV vaccines have largely stalled due to political resistance and budgetary concerns (NCSL 2014). Given the difficulty of implementing strong public health interventions to increase HPV immunization in the U.S., researchers have aimed to identify more incremental policy levers for promoting uptake (Dunne 2014).

Value-based insurance design in U.S. health reform

Value-based insurance design (VBID) is one approach that has been proposed to encourage utilization of cost-effective services. Interventions based on VBID seek to align patients' financial incentives with evidence-based care by selectively modifying patient cost-sharing requirements (Fendrick 2012; Fendrick 2002). Recently, the ACA instated a large-scale VBID intervention through a provision requiring all non-grandfathered commercial health insurance plans to fully cover preventive services rated "A" or "B" by the U.S. Preventive Services Task Force or the ACIP (KFF 2011). The list of covered services includes routine vaccinations and annual wellness visits (during which vaccines may be offered and administered). The removal of cost barriers for HPV vaccination under health reform may incentivize higher rates of vaccine acceptance among parents and age-eligible patients.

The impact of ACA-mandated cost-sharing reductions on HPV immunization is of substantive interest, given the vaccines' potential to prevent most HPV-associated cancers if used optimally. HPV vaccination is also a good test case in that it represents a type of care that has rarely been examined in prior evaluations of VBID: Although existing evidence suggests that patients respond to cost-sharing reductions with respect to essential health services, most of this evidence comes from initiatives by a single employer or insurer targeting prescription drug adherence for chronic conditions (Fendrick 2012). A few studies have looked at the effects of cost-sharing exemptions on the use of preventive services (Rowe 2008; Busch 2006; Reed 2009) but did not focus on pediatric or adolescent populations, which may respond differently to cost-sharing changes.

To examine the effect of vaccine-specific cost-sharing reductions on uptake, I use historical claims data from a sample of commercial health plans that existed during years before and after the ACA (2009-2013). The ACA's VBID intervention gives rise to a natural

experiment among study plans: A portion of plans retained grandfathered status through 2011, 2012, or 2013 and were therefore exempt from the cost-sharing rule for all or part of the post-ACA period; meanwhile, among non-grandfathered plans, the intensity of the VBID intervention varied depending on the baseline (pre-ACA) cost-sharing scenario. Within plans, the study sample includes repeated age-specific cross-sections of still-unvaccinated girls and women in each year from 2009 to 2013. I use Cox proportional hazards models to estimate the effect of cost-sharing reductions on vaccination rates, taking advantage of within-plan variation in cost-sharing levels over time and/or cross-sectional variation in cost-sharing across plans that had similar pre-ACA cost-sharing levels. The models adjust for complex underlying trends in uptake, and include group-level fixed effects to protect against the most likely sources of time-constant omitted variable bias.

METHODS

Policy setting and data source

ACA preventive care provisions

Section 2713 of the ACA requires that any new or non-grandfathered commercial health plan eliminate cost-sharing for recommended preventive services that are provided in-network. There are several caveats to the requirement. For example, plans have the flexibility to use "medical management" techniques such as limiting the types of providers that patients can visit to receive a particular service for free (Dept. of the Treasury 2010 July 19). Additionally, if the provider bills for the office visit and the preventive service separately, cost-sharing may be imposed for the office visit portion (Dept. of the Treasury 2010 July 19).1 Of particular

[1] The intent of this separate billings caveat was to prevent patients from receiving unrelated services for free by asking for a preventive service during the same visit.

relevance for the multi-dose HPV vaccine, federal regulations have only recently clarified that insurers must fully cover any additional well-woman visits required to complete a recommended preventive service (effective for policy years starting on or after August 1, 2012) (Dept. of HHS 2012). Due to these caveats, real-world out-of-pocket costs for included services may not be strictly zero in affected health plans, especially during the early years of ACA implementation.

The preventive care provisions apply starting in the first policy year on or after September 23, 2010 in which a health plan no longer has grandfathered status (which for most plans, would be no earlier than January 1, 2011). Plans may claim grandfathered status if reductions in benefit generosity and employer premium contributions since March 23, 2010 (the ACA's enactment date) do not exceed pre-specified thresholds (Dept. of the Treasury 2010 June 17). Although grandfathered plans are exempt from the rule, some may have voluntarily complied by fully or partially reducing cost-sharing for these services.

Claims data

MarketScan Commercial Claims & Encounters (2007-2013) is a de-identified, individual-level database containing health care claims from over 100 employers and insurers. Although not nationally representative, the database provides a large convenience sample with enrollees in all U.S. states and 33 million covered employees and dependents in 2007. Claims are integrated from all outpatient, inpatient, and pharmacy providers that submitted for reimbursement through the enrollee's health plan. Enrollment files contain basic person-level demographics and plan information, as well as socioeconomic measures based on enrollees' 5-digit ZIP code of residence.

Although the grandfathered status of MarketScan plans is not observable through claims, it is possible to empirically estimate HPV vaccine-related cost-sharing levels by plan and calendar year. For a subset of plans, additional information is available through a supplemental database (the 2010-2013 MarketScan Summary Plan Description Extractions) to help validate empirical cost-sharing estimates. The Extractions database is prepared by Truven Analytics using standardized keyword searches of plan materials, and contains extracted statements relating to preventive care benefits (including vaccinations and wellness visits specifically) and grandfathered status.

Overview of approach

I use a repeated cross-sectional design with plan-level fixed effects to identify the per-dollar impact of cost-sharing reductions on HPV vaccination rates. Within each plan-employer group, my analysis compares vaccination rates among still-unvaccinated 9-26 year-old female beneficiaries in each year of the pre- and post-ACA periods. Cost-sharing policy is modeled as a continuous time-varying variable that is updated once per plan per year. This approach exploits variation in the intensity and timing of cost-sharing reductions across different plan-employer groups to distinguish policy effects from temporal trends in vaccination that are unrelated to the intervention. Plans with relatively little change in cost-sharing levels (e.g., grandfathered plans and plans that already featured zero/low cost-sharing) serve as controls for estimating these background time trends.

Plan-employer fixed effects are included to adjust for any observed or unobserved time-constant group characteristics associated with both HPV vaccine cost-sharing and uptake.

Of particular concern is the possibility of confounding by overall benefit generosity, which is likely to be correlated with vaccine-specific cost-sharing level in a given plan/year. Without adjusting for generosity, two types of confounding are likely to arise: First, individuals with greater propensity to seek health services may tend to select into more generous plans. Second, overall generosity itself could be an important driver of HPV vaccination decisions, an effect that must be disentangled from the effect of vaccine-specific cost-sharing. In the absence of a validated measure of actuarial value in MarketScan, I indirectly adjust for generosity by focusing on within-plan changes in vaccination rates from year to year, with the assumption that plans' overall generosity stays roughly constant during 2009-2013.

For this identification approach to be valid, the timing and extent of preventive care cost-sharing changes should not be correlated with other changes in benefits that could have impacted vaccination rates. Thus, one limitation of my approach is the potential for bias due to changes in generosity that trigger a loss of grandfathered status. However, if reductions in preventive care cost-sharing are concurrent with a substantial loss of other plan benefits, the resulting bias would likely make the policy effect estimate more conservative. The maximum allowable changes for grandfathered status are also stringent[2], so any concurrent changes in a plan's generosity may not have been large enough to appreciably affect the composition or vaccination behavior of beneficiaries in the short term. In sensitivity analyses, I use alternative model specifications with varying levels of covariate adjustment to assess the robustness of the primary specification and to provide an upper limit for the policy effect.

The complexity of background trends in HPV vaccination poses another analytical challenge. Rates of vaccination widely vary across the vaccine-eligible age range (9-26 years), as

[2] For example, a plan can lose grandfathered status due to an increase in any copay requirement by an amount exceeding the greater of: (i) 15 percentage points above medical inflation; or (ii) $5 increased by medical inflation (Dept. of the Treasury 2010 June 17).

HPV vaccines are primarily targeted to young adolescents prior to sexual debut. Moreover, due catch-up vaccinations among older adolescents and young women in the early years of vaccine availability, calendar year trends in vaccination rates are also likely to be age-specific.[3] Statistical analyses are therefore conducted using semi-parametric Cox proportional hazard models, which conveniently adjust for background trends in vaccination rates that occur along the dimensions of age, calendar time, and cohort (i.e., age/calendar time interactions). The main model's time scale is linked to attained age to account for changes in the hazard of vaccination as a subject ages and remains unvaccinated. Models also include interactions between time-varying dummy indicators for each calendar year and attained age, thereby allowing calendar year trends to vary by age.

Sample selection

Plan-employer groups are included in the main study sample if they continuously contributed data to MarketScan from 2009 until 2011 or later. In sensitivity analyses, I either place further restrictions on the set of included plans (e.g., requiring plans to be continuously present in the database from 2009 through 2013) or expand the sample to all plans existing for any portion of 2009-2013.

Within plans, I identify female beneficiaries who were vaccine-eligible in the study timeframe, including those who: (i) newly reached the age of vaccine eligibility during 2009-2013; or (ii) were older but still unvaccinated as of 2009. The former group includes enrollees

[3] The first HPV vaccine became widely available in 2007. So, for example, unvaccinated 15 year-olds in 2009 have only had a two-year window of opportunity to initiate vaccination. Due to the catch-up vaccination effect, they may have higher vaccine initiation rates than unvaccinated 15 year-olds in 2010, who have already had an extra year to initiate vaccination but did not do so. Conversely, unvaccinated 10 year-olds in 2009 and those in 2010 have had the same length of time to initiate vaccination (i.e., 1 year). Background calendar year trends in vaccination are therefore likely to be different for 9 year-olds than for 15 year-olds.

who were 9 or 10 years old as of January 1 in each year from 2009 to 2013. (When possible, girls are followed from age 9; otherwise, they are followed from age 10 and assumed to have not been vaccinated previously, since vaccine initiation rates were found to be very low among 9 year-old females in MarketScan.) The latter group includes enrollees 11-26 years old as of January 1, 2009 who had been continuously enrolled in the MarketScan database since January 1, 2007 (or at least since January 1, 2008 for 11-year-olds), and had not initiated the HPV vaccine series during that time.

By following these individuals over time until vaccination or censoring, I obtain repeated cross-sections of unvaccinated girls and women at each vaccine-eligible age for each year from 2009 to 2013.

Variable measurement

Estimates of plan-level cost-sharing by year

Among affected plans, the intensity of the VBID policy effect varies depending on the pre-ACA cost-sharing scenario. The main analysis therefore uses a continuous measure of cost-sharing level that varies by calendar year in each plan. Alternative functional forms of the policy variable are also considered in sensitivity analyses.

To empirically estimate HPV vaccine-related cost-sharing levels, I extract all in-network outpatient claims associated with HPV vaccine procedure codes in which the patient is female and 9-26 years old on the service date. I then identify claims for outpatient visits in which the vaccine was administered, including any outpatient record on the same date with procedure codes for vaccine administration and/or well-child or well-woman preventive care visits (Appendix A.1). All forms of cost-sharing (copay, coinsurance, and/or deductible) are summed for each vaccination event (i.e., per dose), and then averaged by plan/year. Amounts are

inflation-adjusted to 2013 USD using the medical care component of the Consumer Price Index. For the regression analyses, cost-sharing amounts are divided by 10 to obtain effect estimates per $10 change in cost-sharing level.

Large changes in vaccine cost-sharing requirements were not expected during the pre-ACA period. Therefore, to assess the precision of empirical cost-sharing estimates, I analyze the concordance between cost-sharing estimates for the same plans in 2009 versus 2010. Based on scatterplots and correlation tests, the concordance between cost-sharing estimates is high among study plans (sample size-weighted correlation=0.912), especially in large plan-employer groups that contribute the bulk of the individual-level sample (Appendix A.2). These tests do not confirm concordance across plans with the same true cost-sharing level; however, the primary regression approach only relies on absolute within-plan differences in cost-sharing across calendar years.

Additional validations are performed using the subset of study plans that appear in the Extractions database for any policy year(s) during 2010-2013 (Appendix A.3). Empirical cost-sharing estimates are consistently low in plans that stated zero cost-sharing for routine vaccinations and annual wellness visits (under $5 per dosage for the average plan; ≤$10 for 96% of plans). Estimates are generally higher in plans that stated cost-sharing requirements (mean: $27.75; percent ≤$10: 44%), especially in plans with incomplete coverage of both vaccination itself and associated visits (mean: $94.40; percent ≤$10: 25%). In post-ACA years, there is also strong differentiation between plans that did versus did not claim grandfathered status in plan materials.

Time to HPV vaccine doses

The primary endpoint of interest is time to vaccine initiation (dose 1), measured in years from age-qualification (i.e., time zero ≡ January 1st when the beneficiary was 9 years old). However, subjects have delayed entry into the risk set if they reached age 9 before 2009, or if they were followed from age 10. For example, an 11 year-old in 2009 would enter the risk set at the two-year mark. By linking the time scale of the survival analysis to attained age rather than calendar time, it is possible to estimate the policy effect using within-plan variation across calendar years (i.e., by comparing vaccine initiation rates between plan beneficiaries with the same attained age in different years). In a sensitivity analysis, I instead measure time to vaccine initiation in years from January 1, 2009; under this model, enrollees in the same plan face the same cost-sharing level at every time *t*, and the cost-sharing effect can only be estimated using across-plan variation.

Secondary endpoints include times from age-qualification to receipt of the second and third vaccine dose. In order to directly examine the influence of cost-sharing changes on multi-dose compliance, I also analyze times to two-dose and three-dose completion in terms of months from the preceding dose (conditional on receiving the preceding dose).

In all analyses, subjects are censored at the earliest of: (i) end of continuous enrollment in MarketScan; (ii) December 31, 2013; or (iii) December 31st for those at age 26 years.

Individual- and plan-level covariates

Models adjust for beneficiary and plan characteristics likely to be correlated with both cost-sharing level and background vaccination rates. Time-constant individual-level covariates include census division of residence and the following socioeconomic indicators corresponding to 5-digit ZIP code, derived using American Community Survey 5-year estimates: quartiles for

percentage white race/ethnicity; quartiles for percentage of adult residents without a high school education; and categories of median 4-person family income (≤200%, 201-300%, 301-400%, or >400% federal poverty level for a 4-person family). In subgroup analyses by attained age, I also adjust for employee (versus dependent or spouse) status among beneficiaries with an attained age ≥18 years. In specifications without plan-employer fixed effects, I include covariates for health plan type and data contributor type (insurer or large employer), which are time-constant at the plan level.

Statistical analysis

Primary specifications

The primary set of analyses use Cox proportional hazards models with fixed effects for plan-employer group. In Cox models, fixed effects are implemented by stratifying the baseline hazard function. The models assume that each stratum has a unique baseline hazard function (i.e., different background rates of HPV vaccine uptake), but constrain hazard ratios to be equal for all groups. Thus, the hazard rates associated with two different cost-sharing scenarios are assumed to be proportional within, but not necessarily across, strata.

Time to HPV vaccine receipt

Let $h_{ip}(t)$ be the hazard rate of HPV vaccine initiation for individual i in plan p at time t years from age-qualification. The hazard at time t refers to the instantaneous rate of vaccine initiation (i.e., initiations per person-years) that applies during the age interval $(9+t, 9+t+1]$ for a still-unvaccinated beneficiary. It is approximately equal to the conditional probability of

vaccination by age 9+t+1 given that a beneficiary is still unvaccinated at age 9+t. With plan-level fixed effects, $h_{ip}(t)$ is modeled by the following regression:

$$h_{ip}(t) = h_{0p}(t)\exp\{\beta_1 cs_{ip}(t) + \boldsymbol{\beta_2}'\mathbf{Z_i} + \boldsymbol{\alpha}'\mathbf{Age\_Year_i(t)}\}$$

$$h_{ip}(t) = h_{0p}(t)\exp\left\{\beta_1 cs_{ip}(t) + \boldsymbol{\beta_2}'\mathbf{Z_i} + \sum_{y=2009}^{2012}\sum_{a=9}^{26}\alpha_{ya}year_{yi}(t) * age_a(t)\right\}$$

In this equation, $h_{0p}(t)$ is the baseline hazard function for plan p, corresponding to a beneficiary with zero or reference values for all covariates. The policy variable, $cs_{ip}(t)$, is a time-varying continuous variable for HPV vaccine cost-sharing that individual i faces in plan p at time t. There is no variation in the values of $cs_{ip}(t)$ among plan p enrollees who reach time t in the same calendar year (i.e., those from the same birth year cohort). $\mathbf{Z_i}$ is the set of time-constant variables for individual- and ZIP code-level characteristics, while $\mathbf{Age\_Year}_i(t)$ denotes a series of interaction terms to control for age-specific trends in vaccine initiation across calendar years (e.g., 9-year-olds in 2009, 2010, 2011 or 2012 vs. 2013). (Because the risk set at a given time t only includes vaccine-eligible individuals with the same attained age, there is no need to include age as a main effect in the model.)

This formulation non-parametrically controls for background vaccine initiation rates by attained age within each plan. The plan-specific baseline hazard function absorbs all observed and unobserved plan-level factors that affect rates of vaccine initiation by age, provided that they remain constant over calendar time. Meanwhile, the linear component of the model specifies that, in the absence of cost-sharing or other covariate changes, all plans would have the same proportional changes in age-specific hazard rates across different calendar years. Similarly, plans face the same proportional change in hazard rates per $10 change in cost-sharing. Inference is based on relative within-plan changes in age-specific hazard rates from year to year in plans

subject to varying intensities of the VBID policy. To the extent that a plan maintains steady cost-sharing levels during the study timeframe, it serves as a control group for estimating background time trends.

The parameter of interest is the hazard ratio of vaccine initiation per $10 reduction in cost-sharing, equal to $1/\exp\{B_1\}$. Thus, $(1/\exp\{B_1\} - 1)$ gives the percentage change in the hazard of vaccine initiation per $10 reduction. To facilitate comparisons with prior studies that estimated the price sensitivity of demand for health services, I also estimate the price elasticity of demand for the HPV vaccine, i.e., the percentage change in the hazard rate per 1% change in price. For elasticity calculations, I use the average vaccine-related cost-sharing levels in 2009 and 2013 as the starting and ending price points, respectively.

Analogous methods are used to estimate the policy effect on times from age-qualification to receipt of the second and third vaccine doses.

Time gap between doses (conditional models)

The conditional models of two-dose and three-dose compliance are estimated among beneficiaries who had received the previous dose in the HPV vaccine series. The time origin in these analyses is the previous dose date.

Let $h_{ip}(t^*)$ be the hazard rate of receiving the second (third) dose for individual *i* in plan *p* at time t^* months from receiving the first (second) dose. With plan-level fixed effects,

$$h_{ipc}(t^*) = h_{0c}(t^*)\exp\{\beta_1 cs_{ip}(t^*) + \boldsymbol{\beta_2}'\mathbf{Z_i} + \boldsymbol{\alpha}'\mathbf{Age\_PriorDose_i} + \boldsymbol{\gamma}'\mathbf{Year\_PriorDose_i}\}$$

$$h_{ipc}(t^*) = h_{0c}(t^*)\exp\left\{\beta_1 cs_{ip}(t^*) + \boldsymbol{\beta_2}'\mathbf{Z_i} + \sum_{a=9}^{25}\alpha_a \text{age\_priordose}_{ai} + \sum_{y=2009}^{2012}\gamma_y \text{year\_priordose}_{yi}\right\}$$

where $h_{0p}(t^*)$ is the baseline hazard function for plan *p*, and $\mathbf{Age\_PriorDose}_i$ and $\mathbf{Year\_PriorDose}_i$ are vectors of time-constant dummy variables indicating individual *i*'s age and calendar year at the time of receiving the previous dose. The variables $cs_{ip}(t^*)$ and $\mathbf{Z}_i$, are defined as before.

Sensitivity testing

Testing non-linear functions of cost-sharing

The proportional hazard models include vaccine-related cost-sharing as a linear predictor. This functional form assumes a constant percentage increase in vaccination rates per dollar reduction regardless of the starting price, implying a larger percentage increase in rates per 1% reduction in price at higher starting prices. There is empirical support for this assumption, as studies have consistently found greater price elasticity of health care demand at higher cost-sharing levels (Newhouse & Phelps 1974; Rosett & Huang 1973; Newhouse 1993). However, I also assess model fit using nonlinear functions of cost-sharing, including a log-transformation of cost-sharing plus $1 (which assumes a more stable price elasticity of demand across the range of prices) or a quadratic function of cost-sharing (which can either stabilize or further increase price elasticity at higher price points). Additionally, I test a dichotomous policy variable ("free/near-free" vs. "not free") defined by cost-sharing level below or above a low threshold ($10); as discussed in Appendix A.4, this specification yields a very conservative estimate of the overall VBID policy effect among study plans.

Based on changes in the -2 log L statistic (Collett 2003), model fit does not significantly improve when a squared cost-sharing term is added to the model, and worsens when the linear variable is replaced with a log-transformation or binary cost-sharing variable (Appendix A.4). I

therefore conclude that demand for HPV vaccination is adequately modeled using a linear price variable and use this functional form throughout the analyses.

Alternative model specifications

Alternative specifications with varying levels of covariate adjustment are tested to examine the potential range of effect sizes and to assess robustness. The most parsimonious model includes age*year covariates only, without group-level fixed effects; this naïve model arguably provides an upper limit for the per-dollar impact of cost-sharing reductions, as it uses cross-sectional variation in vaccine-related cost-sharing across plans with potentially very different generosity levels and beneficiary characteristics. To gauge the potential for omitted variable bias in the main specification, I compare the incremental effect of adding individual-level covariates to the naïve model versus adding them to a model containing age*year covariates and plan-level fixed effects only. If adjusting for important individual-level confounders has little impact when combined with plan-level fixed effects, it would mitigate concerns about time-varying confounding due to within-plan changes in beneficiary composition.

Lastly, I try a different fixed effects strategy with stratification by 2009 cost-sharing level *c* ($0-10, >$10-25, >$25-50, or >$50), with the assumption that differences in overall benefit generosity are negligible among plans with similar pre-ACA cost-sharing. For example, the regression model for vaccine initiation is:

$$h_{ipc}(t) = h_{0c}(t)\exp\{\beta_1 cs_{ip}(t) + \boldsymbol{\beta_2}'\mathbf{Z_i} + \boldsymbol{\beta_3}'\mathbf{X_p} + \boldsymbol{\alpha}'\mathbf{Age\_Year_i(t)}\}$$

$$h_{ipc}(t) = h_{0c}(t)\exp\left\{\beta_1 cs_{ip}(t) + \boldsymbol{\beta_2}'\mathbf{Z_i} + \boldsymbol{\beta_3}'\mathbf{X_p} + \sum_{y=2009}^{2012}\sum_{a=9}^{26}\alpha_{ya}\,year_{yi}(t) * age_a(t)\right\}$$

where $h_{ipc}(t)$ is the hazard rate of HPV vaccine initiation for individual i in plan p with pre-ACA cost-sharing level c at time t years from age-qualification, and $\mathbf{X_p}$ is the set of time-constant plan-level covariates. This model utilizes both within-plan variation in cost-sharing over time and across-plan variation within each 2009 cost-sharing stratum.

Sensitivity analyses using different sampling schemes

I re-estimate the primary regression model using a variety of alternative plan selection criteria to examine potential information and attrition biases. To test the impact of measurement error in the policy variable, I restrict the analysis to large plan-employer groups, in which cost-sharing levels appear to be precisely estimated (Appendix A.2). To test the impact of measurement error with respect to vaccine receipt, I separately estimate the policy effect in groups with ≤ or >$50 cost-sharing in 2009. Unobserved vaccinations should be infrequent in this privately-insured population, but could be more common among beneficiaries facing the full out-of-pocket cost of HPV vaccination: Children up to 18 years old may receive free vaccines through the federal Vaccines For Children (VFC) program if their insurance excludes coverage, although there are important caveats to VFC eligibility for the underinsured (see Discussion section). Lastly, I explore the impact of plan attrition from the study sample by focusing on plans existing in MarketScan for ≥4 years (2009-2012+) or all 5 years (2009-2013), and by broadly including all MarketScan plans present for any part of 2009-2013. (The latter analysis is conducted both with and without plan-level fixed effects, given the lack of multi-year continuity for some plans in the expanded sample.)

Alternative study design

In an additional sensitivity test, I use an alternative longitudinal study design in which the time scale of the survival analysis is linked to calendar time rather than attained age. Unvaccinated women 9-26 years old in 2009 are followed from January 1, 2009 (the new time origin) until vaccine receipt or censoring from the database. In this way, all subjects enter the risk set at the time origin, rather than some having delayed entry depending on starting age in 2009.

The Cox regression models of years from January 2009 until vaccine receipt are stratified by each unique combination of starting age in 2009 and pre-ACA cost-sharing category ($0-10, >$10-25, >$25-50, or >$50). The hazard ratio associated with cost-sharing changes is estimated solely using cross-sectional variation in cost-sharing among plans with similar baseline cost-sharing. As noted earlier, this alternative design is not compatible with plan-level fixed effects, since the risk set within a plan faces the same cost-sharing level at every time step. However, this specification is useful because it varies the common trends assumption in the main specification. Namely, hazards are allowed to vary non-parametrically by attained age*calendar time within each grouping of plans. Thus, background changes in vaccination rates from year to year are no longer assumed to be the same across all plans.

Subgroup analyses

To separately estimate the policy effect within different age ranges, I re-run the primary model of time to vaccine initiation with interactions between $cs_{ip}(t)$ and dummy indicators for categories of attained age at time t. I similarly examine heterogeneity in beneficiaries' sensitivity to the cost-sharing changes across ZIP code-level socioeconomic subgroups. This set of analyses will suggest the likely impact of the VBID intervention on reported disparities in HPV vaccine

uptake (initiation and/or completion) by income, education, and race/ethnicity (Reagan-Steiner 2015; Rahman 2013). I conduct socioeconomic subgroup analyses for both vaccine initiation and multi-dose compliance, given evidence that disparities by race/ethnicity mainly occur with respect to vaccine completion among those who initiate.

RESULTS

Descriptive statistics

Beneficiary and plan characteristics

The primary study sample includes 1,680,608 unique beneficiaries from 1,764 plan-employer groups, with a total of 3,950,885 person-years of enrollment across 2009-2013 (Tables 1.1-1.2). The majority of subjects were enrolled in plans featuring low or moderate in-network cost-sharing for HPV vaccines in 2009 (78% with cost-sharing ≤$25 per dosage). However, 3.1% of the individual-level sample (52,168 beneficiaries with 122,362 person-years of enrollment) came from plans with cost-sharing levels over $50 per dosage in 2009.

On average, vaccine-eligible girls and young women were 13.3 years old in their first year of follow-up during 2009-2013. Plan-employer groups with varying levels of pre-ACA cost-sharing had comparable age distributions of beneficiaries, but differed substantially in terms of geographic distribution and plan characteristics. For example, plans in the lower cost-sharing stratums were more likely to be health maintenance organization types and less likely to be preferred provider organization types. Beneficiaries in plans with more generous pre-ACA vaccine coverage also tended to reside in ZIP codes with higher median family income.

Table 1.1: Beneficiary characteristics

	All study plans	By pre-ACA (2009) cost-sharing			
		$0-10	>$10-25	>$25-50	>$50
N	1,680,608	500,348	815,787	312,305	52,168
Age at first year in risk set					
Mean (SD)	13.3 (5.0)	13.3 (4.9)	13.2 (5.0)	13.3 (5.1)	13.6 (5.2)
Age group, %					
9-12 years	59.0	57.9	59.7	59.3	56.5
13-17 years	19.0	19.3	19.0	18.0	20.3
18-26 years	22.1	22.9	21.2	22.7	23.2
Census division, %					
East North Central	23.9	19.7	32.1	7.3	35.7
East South Central	5.2	4.7	6.4	2.7	5.6
Middle Atlantic	7.4	13.4	6.2	1.8	3.4
Mountain	5.6	4.5	6.2	5.9	7.0
New England	2.9	4.5	2.7	1.2	0.9
Pacific	10.4	23.4	6.0	2.3	3.2
South Atlantic	16.1	16.8	19.1	8.3	9.9
West North Central	4.6	4.0	5.0	2.7	14.3
West South Central	23.7	8.6	16.3	67.7	20.0
Unknown	0.2	0.4	0.1	0.1	0.0
Insurer data contributor type (vs. employer)[b], %	55.2	26.8	60.9	87.9	42.2
Plan type, %					
Comprehensive	1.1	1.0	1.2	0.6	3.7
Exclusive provider organization	1.0	2.8	0.3	0.0	0.0
Health maintenance organization	13.3	25.9	10.4	2.9	0.5
Point-of-service (non-capitated)	8.4	9.6	6.1	13.8	0.7
Preferred provider organization	70.1	53.4	76.7	76.7	88.8
Point-of-service (with capitation)	0.4	0.4	0.4	0.1	0.0
Consumer-driven health plan	2.5	5.1	1.6	1.2	0.5
High deductible health plan	1.3	1.4	1.1	1.8	1.3
Unknown/other	1.8	0.3	2.1	3.0	4.6
SES measures by zip code[c], %					
Median 4-person family income					
≤200% FPL	10.2	9.2	9.1	14.2	14.5
201-300% FPL	25.9	24.1	25.2	29.4	32.5
301-400% FPL	27.4	27.4	27.9	26.7	24.4
>400% FPL	34.2	37.0	35.7	26.9	26.4
Quartiles, % adults without high school completion					
1 (highest completion rate)	30.3	30.7	32.7	24.1	26.8
2	23.8	25.5	24.5	19.6	21.8
3	23.4	21.6	23.9	24.8	23.9
4	20.3	19.9	16.9	28.8	25.4
Quartiles, % white race					
1 (lowest % white)	19.0	23.0	18.2	15.5	13.5

Table 1.1 (Continued)

2	27.0	26.9	26.3	29.9	20.0
3	28.8	25.2	29.4	32.2	31.3
4	23.0	22.5	24.0	19.6	33.1
Missing socioeconomic measures	2.3	2.4	2.0	2.7	2.1

Type of MarketScan data contributor (insurer vs. employer) is a proxy for enrollment in smaller- versus larger-group health plans, respectively. Categorical socioeconomic measures at the 5-digit zipcode level are based on American Community Survey 5-year estimates.
FPL, federal poverty level; SES, socioeconomic.

Table 1.2: Cost-sharing per HPV vaccine dose in study plans by year (2009-2013)

Year	Number of plans	Subjects enrolled	Average (SD) plan cost-sharing	Percentiles				
				1%	25%	50%	75%	99%
2009	1,764	1,113,701	19.92(20)	0	9	19	24	133
2010	1,764	863,212	19.23(22)	0	10	16	23	163
2011	1,764	821,702	13.59(18)	0	6	11	16	147
2012	1,289	739,959	12.14(19)	0	6	10	12	151
2013	982	412,311	7.54(6)	0	4	6	11	27

Dollar amounts are inflation-adjusted to USD 2013. To compute the cost-sharing statistics shown above, plan-level cost-sharing amounts are weighted by the number of subjects enrolled in the specified year. Above, the number of enrollees includes all years of enrollment among study subjects, before or after the vaccination events of interest. Subsequent tables report person-years at risk, which does not include years after an individual has the event.

Trends in vaccine cost-sharing & uptake

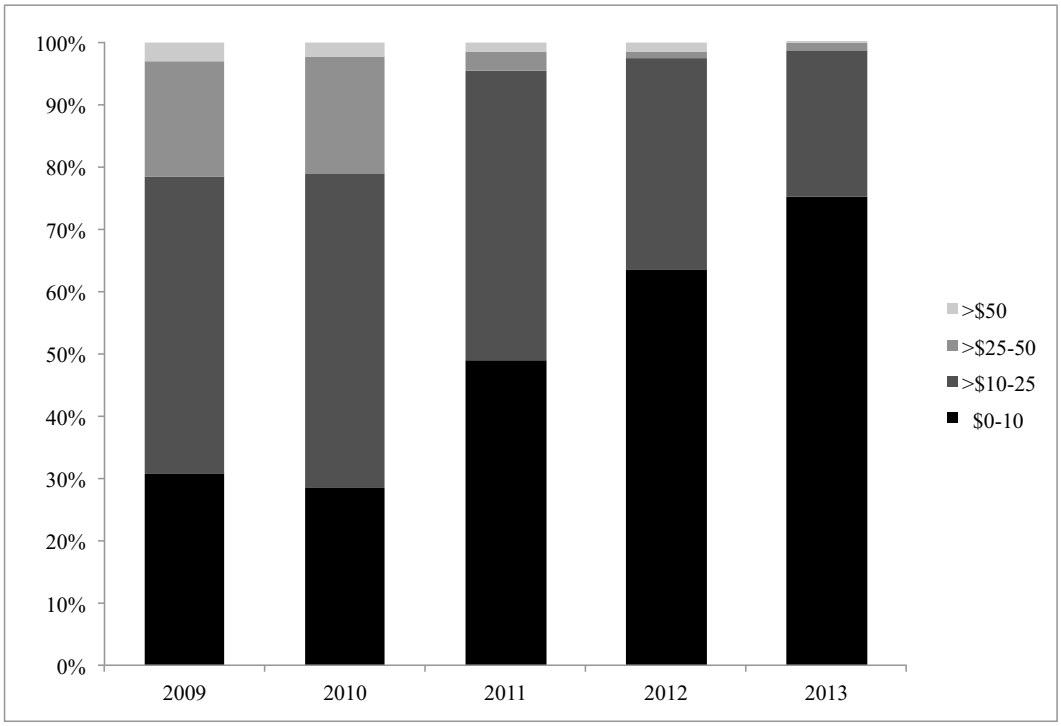
Table 1.2 and Figures 1.1-1.2 describe trends in HPV vaccine-related cost-sharing during the study timeframe. Average cost-sharing across all plans decreased from $19.92 in 2009 to $7.53 in 2013, a -62% (-$12.38) change. The proportion of plans featuring low or no cost-sharing for HPV vaccination ($0-10 per dosage) increased from 31% in 2009 to 75% in 2013. Within all strata of pre-ACA cost-sharing, cost-sharing levels declined during the post-ACA period (2011-2013) after being relatively steady in pre-ACA years (2009-2010). The decreases in cost-sharing are larger in strata with higher pre-ACA cost-sharing levels.

Unadjusted trends in HPV vaccine initiation by year (2009-2013) are plotted in Figure 1.3 for repeated cross-sections of still-unvaccinated women, stratified by attained age range. The

hazard probability of initiation decreased from 2009 to 2010 for all ages above 9 or 10 years old, consistent with an ongoing catch-up trend among girls and women who were past the minimum eligible age at vaccine introduction. Nevertheless, vaccine initiations increased in the post-ACA period for most age groups, and exceeded 2009 levels by 2013 for all ages ≤17 years.

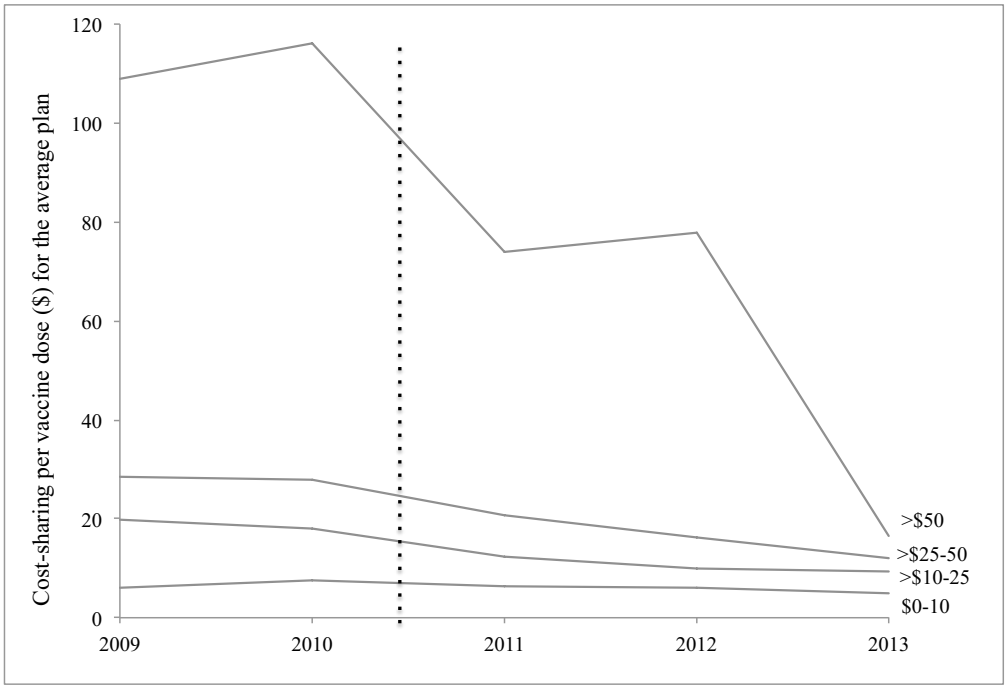
Figure 1.4 plots changes in age-adjusted vaccine initiation rates over time within each pre-ACA cost-sharing stratum (% change vs. 2009, by year). This plot illustrates the divergence in yearly vaccination trends between groups of plans that, on average, were subject to varying intensities of the VBID policy (As seen in Figure 1.2, the VBID policy had more "bite" in plans with higher baseline cost-sharing levels.) The groups had similar decreases in vaccination rates from 2009 to 2010, but groups with higher baseline cost-sharing generally had larger increases in vaccination in post-ACA years (with the exception of the >$25-50 group in 2011 and 2012).

Figure 1.1: Distribution of cost-sharing per HPV vaccine dose in study plans by year



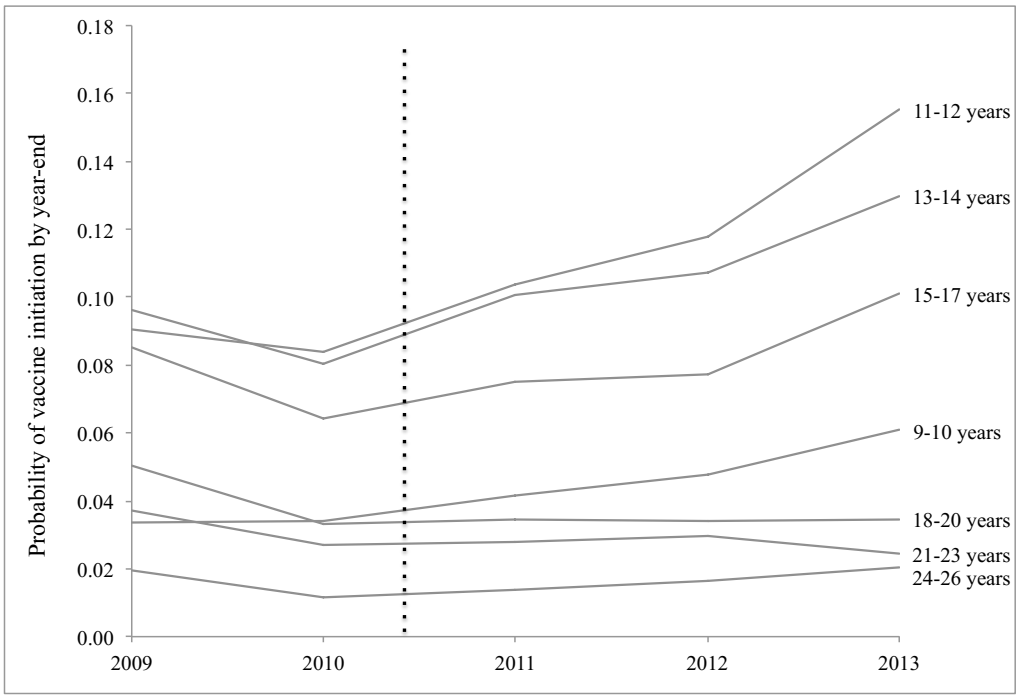
To compute cost-sharing statistics, plans are weighted by the number of study subjects enrolled in the specified year.

Figure 1.2: Average cost-sharing for HPV vaccination among study plans (2009-2013), stratified by 2009 cost-sharing level



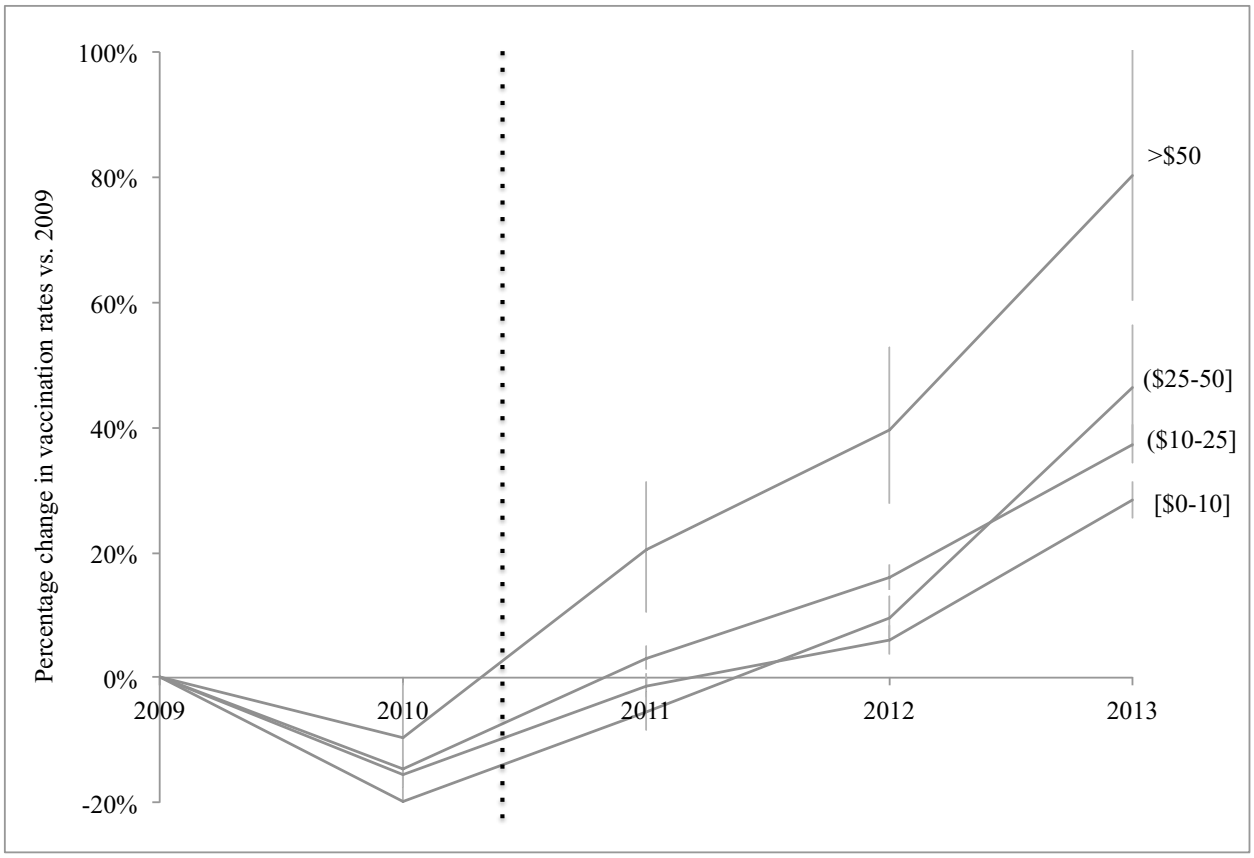
To compute cost-sharing statistics, plans are weighted by number of study subjects enrolled in the specified year.

Figure 1.3: Unadjusted 1-year probabilities of HPV vaccine initiation in repeated cross-sections of unvaccinated subjects, by category of attained age



Each line corresponds to repeated cross-sections of still-unvaccinated beneficiaries with the same attained age range but in different calendar years.

Figure 1.4: Changes in HPV vaccine initiation rates from 2009, stratified by 2009 cost-sharing level



Error bars represent 95% confidence intervals. In this figure, pre-ACA cost-sharing level serves as a proxy for the average intensity of the vaccine-related cost-sharing reductions. For each pre-ACA cost-sharing stratum, the within-stratum changes in vaccine initiation rates are calculated using a Cox model with dummies for each calendar year only. This descriptive model implicitly adjusts for any within-stratum imbalances in age across calendar years, since the risk set at a given time *t* only contains beneficiaries with the same attained age (i.e., age 9+*t* years).

Multivariate results

Policy effect on overall vaccination rates

The main specification with plan-employer fixed effects estimates a 4.1% (95% CI: 3.3-4.9%) increase in the hazard rate of vaccine initiation per $10 reduction in cost-sharing from 2009 ($p<0.001$) (Table 1.3). Effect sizes are similar in corresponding models of time from age-qualification to receipt of the second dose (3.6%; 95% CI: 2.6-4.6%) and third dose (4.3%; 95% CI: 2.9-5.6%) (both $p<0.001$). Coefficients on the age*year covariates indicate sizable increases

in vaccination rates during the post-ACA period, beyond the increases explained by variation in intensity of the VBID policy among study plans.

Based on the same models, Figure 1.5 presents adjusted cumulative vaccine initiation and completion by attained age under various cost-sharing scenarios ($100, $50, $25, or $0 per vaccine dose). The curves are predicted based on average beneficiary characteristics and 2013 vaccination rates for each attained age. By the end of vaccine eligibility at age 27 years, cumulative vaccine initiation/completion under each scenario is 68%/44% (free vaccination), 64%/41% ($25/dose), 61%/38% ($50/dose), and 54%/32% ($100/dose). Median age at vaccine initiation ranges from 16 years under free vaccination to 22 years under $100 cost-sharing.

The finding of a statistically significant policy effect is robust to changes in the model specification, plan selection criteria, and study design (Tables 1.4-1.6). In alternative specifications with varying levels of covariate adjustment, the percent increase in vaccine initiation rates per $10 cost-sharing reduction ranges from 7.5% (naïve model with age*year covariates only) to 4.2% (plan fixed effects model with age*year covariates only) (Table 1.4a). Adding time-constant plan- and individual-level covariates to the naïve model reduces the magnitude of the effect size by 28% (columns [iv] vs. [v]); however, individual-level covariates have little impact when added to a model that already contains plan-employer fixed effects (columns [i] vs. [ii]). The pattern of results is similar in analogous specifications of time from age-qualification to the second dose and third dose (Table 1.4b-c).

In sensitivity analyses using different plan selection criteria (Table 1.5), the effect sizes only slightly increase when restricting the sample to large plans only (e.g., HR of vaccine initiation per $10 reduction: 1.041 vs. 1.050; row [i] vs. [ii]). Effect estimates also do not appreciably change when focusing on plans with pre-ACA cost-sharing levels ≤$50 or >$50

(rows [iii] and [iv]), or when requiring plans to have ≥4 years or all 5 years of data from 2009 (rows [v] and [vi]). In the expanded sample of all MarketScan plans present for any part of 2009-2013, the HR of vaccine initiation per $10 cost-sharing reduction is 1.042 (95% CI: 1.036-1.048) with plan fixed effects and 1.057 (95% CI: 1.054-1.059) without fixed effects (rows [vii] and [viii]), similar to corresponding analyses of the main study sample (Table 1.4, columns [i] and [iv]).

Lastly, in the sensitivity analysis using an alternative study design (Table 1.6), which counts time to each vaccine dose in years from January 1, 2009, the HR (95% CI) per $10 cost-sharing reduction is 1.054 (1.049-1.060) for vaccine initiation. This estimate is comparable to the HR of 1.048 (1.043-1.052) obtained from the primary study design when using fixed effects for pre-ACA cost-sharing strata (Table 1.4, column [iii]).

Table 1.7 summarizes effect sizes and price elasticities of demand for HPV vaccine doses, based on the overall change in vaccine-related cost-sharing levels among study plans. The average cost-sharing reduction of 62% ($12.38 per dose) corresponds to a 5.0% increase in vaccine initiation rates among still-unvaccinated beneficiaries. The price elasticity of demand for the first HPV vaccine dose is estimated at -0.08 (95% CI: -0.07, -0.10; sensitivity analysis range: -0.07 to -0.15), implying that a 10% decrease in price would generate a 0.8% increase in vaccine initiation rates among the unvaccinated.

Table 1.3: Plan-employer fixed effects specifications: Impact of cost-sharing reductions on HPV vaccination rates

Dependent variables: Years from age-qualification to receipt of vaccine doses

Covariate	Dose 1 (Initiation)	Dose 2	Dose 3 (Completion)
	HR (95% CI)	HR (95% CI)	HR (95% CI)
Cost-sharing level at time *t* (per $10 reduction)	1.041 (1.033, 1.048)***	1.036 (1.027, 1.046)***	1.043 (1.029, 1.057)***
Census division			
East North Central	1.20 (1.18, 1.23)***	1.25 (1.22, 1.28)***	1.25 (1.20, 1.29)***
East South Central	1.11 (1.08, 1.14)***	1.10 (1.07, 1.14)***	1.11 (1.06, 1.17)***
Middle Atlantic	1.22 (1.19, 1.26)***	1.28 (1.25, 1.32)***	1.32 (1.27, 1.37)***
Mountain	1.19 (1.16, 1.23)***	1.15 (1.10, 1.19)***	1.07 (1.02, 1.13)**
New England	1.37 (1.32, 1.42)***	1.42 (1.36, 1.48)***	1.46 (1.38, 1.54)***
Pacific	1.41 (1.37, 1.45)***	1.39 (1.34, 1.43)***	1.35 (1.30, 1.41)***
South Atlantic	1.12 (1.09, 1.14)***	1.13 (1.10, 1.16)***	1.13 (1.09, 1.17)***
Unknown	1.07 (0.95, 1.21)	1.12 (0.97, 1.30)	1.03 (0.84, 1.26)
West North Central	1.31 (1.27, 1.35)***	1.35 (1.30, 1.39)***	1.37 (1.31, 1.43)***
West South Central	1.00 (ref)	1.00 (ref)	1.00 (ref)
ZIP code SES factors			
Median family income (4-person)			
≤200% FPL	0.89 (0.87, 0.90)***	0.80 (0.78, 0.82)***	0.73 (0.71, 0.76)***
201-300% FPL	0.88 (0.87, 0.90)***	0.83 (0.82, 0.85)***	0.80 (0.78, 0.82)***
301-400% FPL	0.94 (0.92, 0.95)***	0.91 (0.89, 0.92)***	0.88 (0.87, 0.90)***
>400% FPL	1.00 (ref)	1.00 (ref)	1.00 (ref)
Quartiles, % adults without high school completion			
1 (highest completion rate)	1.11 (1.09, 1.13)***	1.17 (1.15, 1.20)***	1.24 (1.20, 1.28)***
2	1.03 (1.01, 1.05)***	1.06 (1.04, 1.09)***	1.10 (1.07, 1.14)***
3	0.99 (0.97, 1.00)	1.01 (0.99, 1.02)	1.02 (1.00, 1.05)
4	1.00 (ref)	1.00 (ref)	1.00 (ref)
Quartiles, % white race			
1 (lowest % white)	1.11 (1.09, 1.13)***	1.05 (1.03, 1.07)***	1.04 (1.01, 1.06)**
2	1.12 (1.10, 1.13)***	1.09 (1.08, 1.11)***	1.09 (1.07, 1.12)***
3	1.09 (1.08, 1.10)***	1.08 (1.07, 1.10)***	1.09 (1.07, 1.11)***
4	1.00 (ref)	1.00 (ref)	1.00 (ref)
Missing socioeconomic measures	1.02 (0.98, 1.06)	0.97 (0.92, 1.01)	0.95 (0.90, 1.01)
Current age x year			
9 years x 2009	1.01 (0.91, 1.13)	0.97 (0.82, 1.15)	0.96 (0.67, 1.37)
x 2010	0.86 (0.77, 0.96)**	0.78 (0.65, 0.93)**	0.70 (0.48, 1.03)
x 2011	0.88 (0.79, 0.99)*	0.86 (0.73, 1.02)	0.86 (0.60, 1.24)
x 2012	0.88 (0.78, 0.98)*	0.87 (0.73, 1.04)	1.05 (0.73, 1.50)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
10 years * 2009	0.55 (0.53, 0.57)***	0.56 (0.53, 0.59)***	0.54 (0.49, 0.61)***
x 2010	0.58 (0.55, 0.60)***	0.58 (0.55, 0.61)***	0.70 (0.63, 0.79)***
x 2011	0.69 (0.67, 0.71)***	0.69 (0.65, 0.72)***	0.72 (0.64, 0.80)***
x 2012	0.80 (0.77, 0.83)***	0.79 (0.75, 0.83)***	0.92 (0.83, 1.02)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)

Table 1.3 (Continued)

11 years x 2009	0.62 (0.60, 0.64)***	0.50 (0.47, 0.52)***	0.24 (0.23, 0.26)***
x 2010	0.58 (0.56, 0.60)***	0.58 (0.56, 0.60)***	0.63 (0.60, 0.67)***
x 2011	0.70 (0.68, 0.73)***	0.71 (0.68, 0.73)***	0.72 (0.68, 0.76)***
x 2012	0.81 (0.79, 0.84)***	0.81 (0.78, 0.84)***	0.86 (0.81, 0.90)***
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
12 years x 2009	0.65 (0.62, 0.68)***	0.45 (0.43, 0.48)***	0.17 (0.15, 0.18)***
x 2010	0.60 (0.57, 0.62)***	0.57 (0.54, 0.60)***	0.61 (0.58, 0.65)***
x 2011	0.73 (0.70, 0.76)***	0.71 (0.68, 0.74)***	0.69 (0.66, 0.73)***
x 2012	0.79 (0.76, 0.82)***	0.81 (0.77, 0.84)***	0.79 (0.75, 0.83)***
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
13 years x 2009	0.76 (0.73, 0.80)***	0.54 (0.51, 0.57)***	0.22 (0.20, 0.24)***
x 2010	0.64 (0.61, 0.68)***	0.59 (0.56, 0.63)***	0.57 (0.53, 0.61)***
x 2011	0.77 (0.73, 0.80)***	0.76 (0.72, 0.80)***	0.71 (0.67, 0.75)***
x 2012	0.84 (0.81, 0.88)***	0.85 (0.81, 0.90)***	0.86 (0.81, 0.91)***
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
14 years x 2009	0.81 (0.77, 0.85)***	0.54 (0.50, 0.57)***	0.20 (0.19, 0.22)***
x 2010	0.69 (0.65, 0.73)***	0.65 (0.61, 0.69)***	0.64 (0.60, 0.69)***
x 2011	0.81 (0.77, 0.86)***	0.77 (0.72, 0.82)***	0.71 (0.66, 0.76)***
x 2012	0.83 (0.78, 0.87)***	0.84 (0.80, 0.89)***	0.83 (0.78, 0.88)***
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
15 years x 2009	0.87 (0.81, 0.92)***	0.58 (0.54, 0.63)***	0.22 (0.20, 0.24)***
x 2010	0.68 (0.64, 0.73)***	0.65 (0.61, 0.70)***	0.66 (0.61, 0.71)***
x 2011	0.79 (0.74, 0.85)***	0.79 (0.74, 0.85)***	0.76 (0.70, 0.82)***
x 2012	0.82 (0.77, 0.88)***	0.86 (0.80, 0.92)***	0.87 (0.81, 0.94)***
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
16 years x 2009	0.97 (0.90, 1.04)	0.70 (0.64, 0.75)***	0.30 (0.27, 0.34)***
x 2010	0.73 (0.67, 0.78)***	0.73 (0.68, 0.79)***	0.76 (0.70, 0.83)***
x 2011	0.82 (0.76, 0.89)***	0.84 (0.78, 0.91)***	0.82 (0.75, 0.89)***
x 2012	0.83 (0.77, 0.90)***	0.89 (0.82, 0.96)**	0.95 (0.87, 1.04)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
17 years x 2009	1.08 (1.01, 1.15)*	0.78 (0.72, 0.84)***	0.38 (0.35, 0.42)***
x 2010	0.80 (0.74, 0.86)***	0.78 (0.72, 0.84)***	0.82 (0.75, 0.90)***
x 2011	0.86 (0.80, 0.93)***	0.86 (0.80, 0.93)***	0.81 (0.74, 0.89)***
x 2012	0.92 (0.85, 0.99)*	0.93 (0.85, 1.01)	0.93 (0.84, 1.02)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
18 years x 2009	1.22 (1.11, 1.34)***	0.76 (0.68, 0.84)***	0.30 (0.26, 0.34)***
x 2010	0.84 (0.76, 0.93)**	0.81 (0.73, 0.90)***	0.90 (0.81, 1.00)*
x 2011	0.89 (0.80, 0.98)*	0.83 (0.75, 0.93)***	0.81 (0.73, 0.91)***
x 2012	0.90 (0.81, 1.00)*	0.90 (0.81, 1.01)	0.95 (0.86, 1.07)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
19 years x 2009	1.60 (1.41, 1.82)***	0.91 (0.80, 1.04)	0.41 (0.35, 0.49)***
x 2010	1.07 (0.94, 1.22)	0.94 (0.82, 1.07)	1.10 (0.94, 1.28)
x 2011	1.05 (0.91, 1.20)	0.94 (0.82, 1.08)	0.90 (0.77, 1.05)
x 2012	1.05 (0.92, 1.21)	0.94 (0.81, 1.08)	0.95 (0.81, 1.12)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
20 years x 2009	1.78 (1.56, 2.03)***	1.16 (1.00, 1.34)*	0.56 (0.47, 0.67)***
x 2010	1.10 (0.96, 1.27)	1.04 (0.89, 1.21)	1.09 (0.92, 1.29)
x 2011	1.05 (0.91, 1.22)	0.97 (0.83, 1.13)	0.84 (0.70, 1.01)
x 2012	1.06 (0.91, 1.23)	0.95 (0.81, 1.12)	0.91 (0.76, 1.09)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
21 years x 2009	1.64 (1.43, 1.88)***	1.05 (0.90, 1.22)	0.59 (0.48, 0.72)***

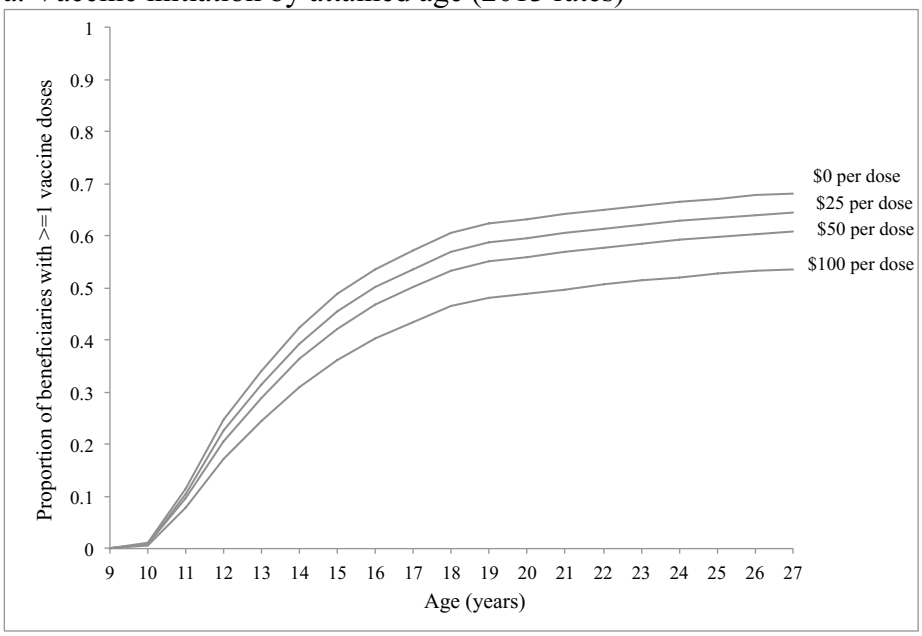
Table 1.3 (Continued)

x 2010	1.15 (0.99, 1.33)	1.13 (0.97, 1.32)	1.38 (1.14, 1.66)***
x 2011	0.97 (0.84, 1.13)	0.94 (0.80, 1.11)	1.18 (0.97, 1.43)
x 2012	0.99 (0.85, 1.16)	0.95 (0.80, 1.12)	0.94 (0.77, 1.16)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
22 years x 2009	1.27 (1.09, 1.47)**	0.91 (0.76, 1.08)	0.48 (0.38, 0.59)***
x 2010	0.96 (0.81, 1.12)	1.03 (0.86, 1.23)	1.09 (0.89, 1.34)
x 2011	1.03 (0.87, 1.21)	1.11 (0.92, 1.33)	1.05 (0.85, 1.29)
x 2012	1.13 (0.96, 1.33)	1.05 (0.87, 1.27)	0.92 (0.74, 1.15)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
23 years x 2009	1.50 (1.24, 1.81)***	0.91 (0.74, 1.13)	0.56 (0.42, 0.73)***
x 2010	1.17 (0.95, 1.44)	0.98 (0.78, 1.22)	1.31 (1.01, 1.70)*
x 2011	1.42 (1.16, 1.73)***	1.18 (0.95, 1.46)	1.10 (0.85, 1.43)
x 2012	1.33 (1.10, 1.61)**	1.13 (0.93, 1.39)	1.09 (0.85, 1.40)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
24 years x 2009	1.20 (0.97, 1.48)	0.96 (0.76, 1.22)	0.50 (0.37, 0.67)***
x 2010	0.76 (0.59, 0.97)*	0.78 (0.59, 1.02)	0.74 (0.53, 1.02)
x 2011	0.98 (0.77, 1.24)	1.05 (0.81, 1.37)	0.87 (0.63, 1.20)
x 2012	1.03 (0.82, 1.30)	1.11 (0.86, 1.42)	1.11 (0.83, 1.48)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
25 years x 2009	1.24 (0.91, 1.69)	0.95 (0.68, 1.34)	0.54 (0.37, 0.79)**
x 2010	0.87 (0.63, 1.22)	1.02 (0.71, 1.46)	1.06 (0.72, 1.55)
x 2011	0.72 (0.50, 1.05)	0.64 (0.43, 0.97)*	0.66 (0.42, 1.04)
x 2012	0.73 (0.51, 1.06)	0.64 (0.43, 0.97)*	0.58 (0.37, 0.91)*
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
26 years x 2009	0.88 (0.35, 2.17)	0.81 (0.29, 2.23)	0.33 (0.13, 0.84)*
x 2010	0.48 (0.19, 1.21)	0.80 (0.29, 2.22)	0.67 (0.27, 1.70)
x 2011	0.50 (0.19, 1.29)	0.79 (0.28, 2.23)	0.74 (0.29, 1.90)
x 2012	0.60 (0.22, 1.66)	0.77 (0.25, 2.36)	0.53 (0.19, 1.51)
x 2013	1.00 (ref)	1.00 (ref)	1.00 (ref)
Plan-employer fixed effects	Y	Y	Y
Pre-ACA cost-sharing fixed effects	N	N	N
Person-years at risk	3,661,077	3,762,297	3,852,750

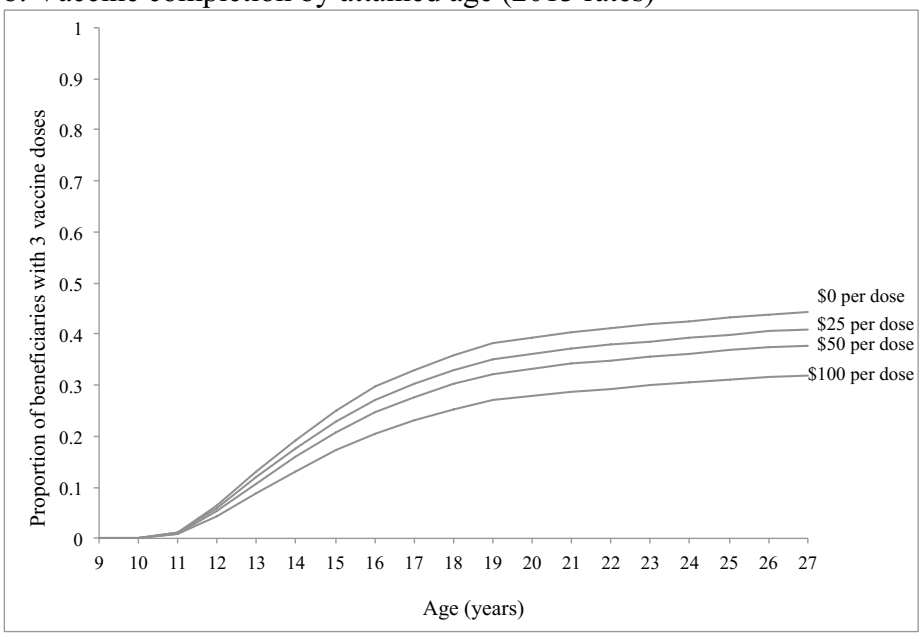
Results are from Cox proportional hazards models with stratification by plan-employer group (fixed effects). All models adjust for attained age through the baseline hazard function. The reference category for the attained age*calendar year interaction terms is the age*2013 because the risk set at a given time *t* only contains beneficiaries with the same attained age.
**p<0.05; **p<0.01; ***p<0.001*

Figure 1.5: Adjusted cumulative initiation and completion of HPV vaccine series under varying cost-sharing scenarios

a. Vaccine initiation by attained age (2013 rates)



b. Vaccine completion by attained age (2013 rates)



Using the primary regression models with plan-level fixed effects, cumulative vaccine initiation and completion by age is predicted under four different cost-sharing scenarios ($0, $25, $50, or $100 per HPV vaccine dose). For each cost-sharing scenario, curves are constructed based on 2013 vaccination rates for each attained age and the average values of time-constant covariates (census division and ZIP code socioeconomic factors) in the overall sample. The curves are first estimated separately for each plan-employer group and then pooled across groups using a weighted average.

Table 1.4: Specifications with varying levels of covariate adjustment: Impact of cost-sharing reductions on overall HPV vaccination rates

a. Dose 1 (Initiation)

Variables	Hazard ratio (95% CI), by model specification				
	(i) Plan fixed effects; all covariates	(ii) Plan fixed effects; age-year covariates only	(iii) Pre-ACA cost-sharing fixed effects; all covariates	(iv) No fixed effects; all covariates	(v) No fixed effects; age-year covariates only
Cost-sharing level at time *t* (per $10 reduction)	1.041*** (1.033, 1.048)	1.042*** (1.034, 1.050)	1.048*** (1.043, 1.052)	1.054*** (1.050, 1.057)	1.075*** (1.072, 1.079)
Age-year trends	Y	Y	Y	Y	Y
Plan-employer fixed effects	Y	Y	N	N	N
Pre-ACA cost-sharing fixed effects	N	N	Y	N	N
Plan and data contributor type	N	N	Y	Y	N
Census division of residence	Y	N	Y	Y	N
ZIP code SES factors	Y	N	Y	Y	N
Person-years at risk	3,661,077	3,661,077	3,661,077	3,661,077	3,661,077

b. Dose 2

Variables	Hazard ratio (95% CI), by model specification				
	(i) Plan fixed effects; all covariates	(ii) Plan fixed effects; age-year covariates only	(iii) Pre-ACA cost-sharing fixed effects; all covariates	(iv) No fixed effects; all covariates	(v) No fixed effects; age-year covariates only
Cost-sharing level at time *t* (per $10 reduction)	1.036*** (1.027, 1.046)	1.039*** (1.029, 1.049)	1.060*** (1.055, 1.066)	1.065*** (1.061, 1.069)	1.088*** (1.084, 1.092)
Age-year trends	Y	Y	Y	Y	Y
Plan-employer fixed effects	Y	Y	N	N	N
Pre-ACA cost-sharing fixed effects	N	N	Y	N	N
Plan and data contributor type	N	N	Y	Y	N
Census division of residence	Y	N	Y	Y	N
ZIP code SES factors	Y	N	Y	Y	N
Person-years at risk	3,762,297	3,762,297	3,762,297	3,762,297	3,762,297

Table 1.4 (Continued)

c. Dose 3 (Completion)

Variables	Hazard ratio (95% CI), by model specification				
	(i) Plan fixed effects; all covariates	(ii) Plan fixed effects; age-year covariates only	(iii) Pre-ACA cost-sharing fixed effects; all covariates	(iv) No fixed effects; all covariates	(v) No fixed effects; age-year covariates only
Cost-sharing level at time *t* (per $10 reduction)	1.043*** (1.029, 1.057)	1.046*** (1.033, 1.060)	1.075*** (1.067, 1.083)	1.075*** (1.069, 1.081)	1.101*** (1.095, 1.107)
Age-year trends	Y	Y	Y	Y	Y
Plan-employer fixed effects	Y	Y	N	N	N
Pre-ACA cost-sharing fixed effects	N	N	Y	N	N
Plan and data contributor type	N	N	Y	Y	N
Census division of residence	Y	N	Y	Y	N
ZIP code SES factors	Y	N	Y	Y	N
Person-years at risk	3,852,750	3,852,750	3,852,750	3,852,750	3,852,750

All models adjust for attained age through the baseline hazard function. Specifications with plan-employer fixed effects do not include time-constant plan-level covariates but implicitly control for these variables.
**p<0.05; **p<0.01; ***p<0.001*

Table 1.5: Sensitivity analyses with different plan selection criteria: Impact of cost-sharing reductions on overall HPV vaccination rates

Plan population		Person-years at risk (Dose 1 model)	Hazard ratio per $10 cost-sharing reduction (95% CI)		
			Dose 1 (Initiation)	Dose 2	Dose 3 (Completion)
Main sample:					
(i)	Plans with 3+ years of data (2009-2011+)	3,661,077	1.041*** (1.033, 1.048)	1.036*** (1.027, 1.046)	1.043*** (1.029, 1.057)
Subsets of main sample:					
(ii)	Plans with N ≥1000 in 2009	2,820,916	1.050*** (1.038, 1.061)	1.046*** (1.032, 1.060)	1.051*** (1.032, 1.070)
(iii)	Plans with cost-sharing ≤$50 in 2009	3,544,302	1.040*** (1.028, 1.053)	1.035*** (1.021, 1.050)	1.039*** (1.019, 1.060)
(iv)	Plans with cost-sharing >$50 in 2009	116,775	1.037*** (1.025, 1.050)	1.031*** (1.015, 1.048)	1.055*** (1.032, 1.079)
(v)	Plans with 4+ years of data (2009-2012+)	3,351,402	1.043*** (1.034, 1.051)	1.040*** (1.030, 1.051)	1.047*** (1.032, 1.062)
(vi)	Plans with 5 years of data (2009-2013)	1,938,232	1.038*** (1.029, 1.048)	1.038*** (1.027, 1.050)	1.039*** (1.023, 1.055)
Expanded sample:					
	Plans with any year(s) of data during 2009-2013:				
(vii)	Plan fixed effects; all covariates	6,807,466	1.042*** (1.036, 1.048)	1.039*** (1.031, 1.047)	1.037*** (1.026, 1.048)
(viii)	No fixed effects; all covariates	6,807,466	1.057*** (1.054, 1.059)	1.067*** (1.064, 1.071)	1.076*** (1.071, 1.081)

In this set of sensitivity analyses, the primary regression specification is re-run using different study plan selection criteria. All Cox proportional hazard models (aside from row [viii]) include plan-level fixed effects and adjust for attained age, age-specific calendar year trends, census division, and ZIP code socioeconomic factors. The plan-level fixed effects are omitted in row [viii] because some plans in the expanded sample are only present in MarketScan for a single calendar year.
**p<0.05; **p<0.01; ***p<0.001*

Table 1.6: Sensitivity analysis using alternative study design: Impact of cost-sharing reductions on overall HPV vaccination rates

Dependent variables: Years from January 2009 to receipt of vaccine doses

Covariate	Dose 1 (Initiation)	Dose 2	Dose 3 (Completion)
	HR (95% CI)	HR (95% CI)	HR (95% CI)
Cost-sharing level at time *t* (per $10 reduction)	1.054 (1.049, 1.060)***	1.069 (1.062, 1.076)***	1.082 (1.073, 1.091)***
Census division			
East North Central	1.18 (1.16, 1.20)***	1.25 (1.22, 1.27)***	1.30 (1.27, 1.33)***
East South Central	1.03 (1.00, 1.06)*	1.07 (1.04, 1.10)***	1.14 (1.10, 1.19)***
Middle Atlantic	1.19 (1.16, 1.22)***	1.26 (1.23, 1.29)***	1.33 (1.29, 1.38)***
Mountain	1.07 (1.05, 1.10)***	1.05 (1.02, 1.09)***	1.03 (0.99, 1.07)
New England	1.50 (1.46, 1.54)***	1.56 (1.51, 1.61)***	1.62 (1.56, 1.69)***
Pacific	1.26 (1.23, 1.28)***	1.21 (1.18, 1.24)***	1.16 (1.13, 1.20)***
South Atlantic	1.01 (1.00, 1.03)	1.04 (1.02, 1.07)***	1.07 (1.04, 1.10)***
Unknown	1.11 (0.99, 1.24)	1.22 (1.07, 1.38)**	1.12 (0.94, 1.35)
West North Central	1.37 (1.34, 1.41)***	1.41 (1.37, 1.46)***	1.49 (1.43, 1.54)***
West South Central	1.00 (ref)	1.00 (ref)	1.00 (ref)
ZIP code SES factors			
Median family income (4-person)			
≤200% FPL	0.82 (0.80, 0.84)***	0.74 (0.71, 0.76)***	0.69 (0.66, 0.71)***
201-300% FPL	0.82 (0.81, 0.84)***	0.77 (0.76, 0.79)***	0.74 (0.72, 0.76)***
301-400% FPL	0.89 (0.88, 0.90)***	0.86 (0.85, 0.87)***	0.84 (0.82, 0.86)***
>400% FPL	1.00 (ref)	1.00 (ref)	1.00 (ref)
Quartiles, % adults without high school completion			
1 (highest completion rate)	1.17 (1.14, 1.19)***	1.24 (1.21, 1.27)***	1.31 (1.27, 1.36)***
2	1.03 (1.01, 1.05)**	1.07 (1.05, 1.09)***	1.12 (1.09, 1.16)***
3	0.98 (0.97, 1.00)	1.00 (0.98, 1.02)	1.03 (1.00, 1.06)
4	1.00 (ref)	1.00 (ref)	1.00 (ref)
Quartiles, % white race			
1 (lowest % white)	1.08 (1.06, 1.10)***	1.01 (0.99, 1.03)	0.99 (0.96, 1.01)
2	1.10 (1.09, 1.12)***	1.08 (1.07, 1.10)***	1.08 (1.06, 1.11)***
3	1.08 (1.07, 1.10)***	1.08 (1.07, 1.10)***	1.09 (1.07, 1.11)***
4	1.00 (ref)	1.00 (ref)	1.00 (ref)
Missing socioeconomic measures	0.97 (0.94, 1.01)	0.92 (0.88, 0.97)***	0.92 (0.87, 0.98)*
Plan-employer fixed effects	N	N	N
Birth year*pre-ACA cost-sharing level fixed effects	Y	Y	Y
Person-years at risk	2,774,097	2,863,854	2,945,977

In this sensitivity analysis, the time origin of the survival analysis is January 1, 2009 (rather than age-qualification for the vaccine). Thus, each model's time scale is linked to calendar time rather than attained age, and all subjects enter the risk set at the time orgin ("time zero"). The sample is restricted to female beneficiaries 9-26 years old in 2009 (N=1,178,016). Models include fixed effects for each unique combination of birth year and pre-ACA cost-sharing level ($0-10, >$10-25, >$25-50, or >$50 in 2009).
**p<0.05; **p<0.01; ***p<0.001*

Table 1.7: Summary of effect sizes and price elasticities of demand for HPV vaccine doses

Endpoint	Percent increase in vaccination rates with $12.38 cost-sharing reduction		Price elasticity of vaccination rates	
	Main estimate (95% CI)	Sensitivity analysis range	Main estimate (95% CI)	Sensitivity analysis range
Dose 1 (initiation)	5.0% (4.1, 6.0)	4.6% to 9.4%	-0.08 (-0.07, -0.10)	-0.07 to -0.15
Dose 2	4.5% (3.3, 5.7)	3.9% to 11.0%	-0.07 (-0.05, -0.09)	-0.06 to -0.18
Dose 3 (completion)	5.4% (3.7, 7.1)	4.6% to 12.6%	-0.09 (-0.06, -0.11)	-0.07 to -0.20

Main estimates are based on the primary model specification with plan-level fixed effects. Elasticity is the percent increase in vaccination rates divided by the percent change in cost-sharing per HPV vaccine dose from 2009 to 2013 (i.e., 62%).

Policy effect on multi-dose compliance

Models with plan-employer fixed effects do not find a significant policy effect on two-dose or three-dose compliance among beneficiaries who had accepted the previous dose (Table 1.8). However, when using less stringent modeling assumptions with either fixed effects for pre-ACA cost-sharing strata or no fixed effects, cost-sharing reductions are associated with small yet statistically significant improvements in multi-dose compliance (Table 1.9a-b, rows [ii] and [iii]). These results are consistent with the models of overall HPV vaccination rates by dose, shown in Table 1.4a-c: In specifications without plan-level fixed effects, effect sizes are slightly larger when the dependent variable is a higher dose level, implying an ongoing effect of cost-sharing reductions beyond its effect on vaccine initiation rates.

Table 1.8: Plan-employer fixed effects specifications (conditional models): Impact of cost-sharing reductions on multi-dose HPV vaccine compliance

Dependent variables: Months between consecutive vaccine doses

Variables	Dose 1 to Dose 2	Dose 2 to Dose 3
	HR (95% CI)	HR (95% CI)
Cost-sharing level at time *t* (per $10 reduction)	1.009 (0.999, 1.019)	1.006 (0.993, 1.020)
Census division		
East North Central	1.08 (1.05, 1.11)***	1.06 (1.03, 1.10)***
East South Central	0.98 (0.95, 1.02)	1.01 (0.97, 1.06)
Middle Atlantic	1.15 (1.11, 1.19)***	1.13 (1.08, 1.18)***
Mountain	0.95 (0.91, 0.99)**	0.90 (0.85, 0.95)***
New England	1.10 (1.06, 1.15)***	1.08 (1.03, 1.15)**
Pacific	1.01 (0.97, 1.04)	0.97 (0.93, 1.02)
South Atlantic	1.03 (1.00, 1.06)*	1.03 (0.99, 1.07)
Unknown	1.04 (0.90, 1.20)	0.96 (0.78, 1.17)
West North Central	1.05 (1.02, 1.09)**	1.05 (1.01, 1.10)*
West South Central	1.00 (ref)	1.00 (ref)
ZIP code SES factors		
Median family income (4-person)		
≤200% FPL	0.85 (0.83, 0.87)***	0.87 (0.84, 0.90)***
201-300% FPL	0.92 (0.90, 0.93)***	0.92 (0.90, 0.95)***
301-400% FPL	0.95 (0.94, 0.97)***	0.96 (0.94, 0.98)***
>400% FPL	1.00 (ref)	1.00 (ref)
Quartiles, % adults without high school completion		
1 (highest completion rate)	1.07 (1.04, 1.09)***	1.09 (1.06, 1.13)***
2	1.04 (1.02, 1.06)***	1.05 (1.02, 1.08)***
3	1.02 (1.00, 1.04)	1.02 (0.99, 1.05)
4	1.00 (ref)	1.00 (ref)
Quartiles, % white race		
1 (lowest % white)	0.89 (0.87, 0.91)***	0.96 (0.93, 0.98)***
2	0.96 (0.94, 0.97)***	0.98 (0.96, 1.00)
3	0.98 (0.97, 0.99)**	0.99 (0.97, 1.01)
4	1.00 (ref)	1.00 (ref)
Missing socioeconomic measures	0.89 (0.85, 0.93)***	0.95 (0.89, 1.00)
Age at preceding dose		
9 years	0.69 (0.62, 0.78)***	0.83 (0.71, 0.96)*
10 years	0.85 (0.76, 0.95)**	0.97 (0.84, 1.12)
11 years	0.83 (0.74, 0.92)***	0.87 (0.75, 1.00)*
12 years	0.83 (0.74, 0.93)***	0.79 (0.69, 0.92)**
13 years	0.83 (0.74, 0.93)**	0.80 (0.70, 0.93)**
14 years	0.83 (0.74, 0.93)***	0.79 (0.68, 0.91)**
15 years	0.83 (0.74, 0.93)**	0.75 (0.65, 0.87)***
16 years	0.84 (0.75, 0.94)**	0.77 (0.66, 0.88)***
17 years	0.76 (0.68, 0.85)***	0.70 (0.60, 0.81)***
18 years	0.70 (0.62, 0.79)***	0.65 (0.56, 0.75)***
19 years	0.74 (0.66, 0.83)***	0.66 (0.57, 0.76)***
20 years	0.76 (0.68, 0.86)***	0.73 (0.62, 0.84)***
21 years	0.76 (0.68, 0.86)***	0.71 (0.61, 0.83)***

Table 1.8 (Continued)

22 years	0.81 (0.72, 0.91)***	0.79 (0.68, 0.92)**
23 years	0.85 (0.75, 0.96)*	0.75 (0.64, 0.88)***
24 years	0.92 (0.81, 1.04)	0.81 (0.68, 0.95)*
25 years	0.96 (0.84, 1.10)	0.80 (0.67, 0.95)*
26 years	1.00 (ref)	1.00 (ref)
Year of preceding dose		
2009	1.05 (1.02, 1.07)***	1.26 (1.22, 1.30)***
2010	1.02 (1.00, 1.04)	1.05 (1.02, 1.09)***
2011	1.02 (1.00, 1.04)*	1.05 (1.02, 1.08)***
2012	0.99 (0.97, 1.01)	1.03 (1.00, 1.06)*
2013	1.00 (ref)	1.00 (ref)
Plan-employer fixed effects	Y	Y
Pre-ACA cost-sharing fixed effects	N	N
N	231,050	156,906
Person-months at risk	1,878,474	1,417,662

Results are from Cox proportional hazard models of months until the next dose in the HPV vaccine series, conditional on receiving the preceding dose.
**p<0.05; **p<0.01; ***p<0.001*

Table 1.9: Specifications with varying levels of covariate adjustment: Impact of cost-sharing reductions on multi-dose HPV vaccine compliance

a. Months from dose 1 to dose 2

Variables	Hazard ratio (95% CI), by model specification		
	(i) Plan fixed effects; all covariates	(ii) Pre-ACA cost-sharing fixed effects; all covariates	(iii) No fixed effects; all covariates
Cost-sharing level at time *t* (per $10 reduction)	1.009 (0.999, 1.019)	1.023*** (1.017, 1.028)	1.022*** (1.018, 1.026)
Age at preceding dose	Y	Y	Y
Year of preceding dose	Y	Y	Y
Plan-employer fixed effects	Y	N	N
Pre-ACA cost-sharing fixed effects	N	Y	N
Plan and data contributor type	N	Y	Y
Census division of residence	Y	Y	Y
ZIP code SES factors	Y	Y	Y
N	231,050	231,050	231,050
Person-months at risk	1,878,474	1,878,474	1,878,474

b. Months from dose 2 to dose 3

Variables	Hazard ratio (95% CI), by model specification		
	(i) Plan fixed effects; all covariates	(ii) Pre-ACA cost-sharing fixed effects; all covariates	(iii) No fixed effects; all covariates
Cost-sharing level at time *t* (per $10 reduction)	1.006 (0.993, 1.020)	1.018*** (1.011, 1.026)	1.013*** (1.007, 1.019)
Age at preceding dose	Y	Y	Y
Year of preceding dose	Y	Y	Y
Plan-employer fixed effects	Y	N	N
Pre-ACA cost-sharing fixed effects	N	Y	N
Plan and data contributor type	N	Y	Y
Census division of residence	Y	Y	Y
ZIP code SES factors	Y	Y	Y
N	156,906	156,906	156,906
Person-months at risk	1,417,662	1,417,662	1,417,662

Results are from Cox proportional hazard models of months until the next dose in the HPV vaccine series, conditional on receiving the preceding dose. Specifications with plan-employer fixed effects did not include time-constant plan-level covariates but implicitly control for these variables.
**p<0.05; **p<0.01; ***p<0.001*

Subgroup analyses

In subgroup analyses by attained age, the cost-sharing reductions are estimated to have a stronger effect on vaccine initiation rates among adolescent girls than young women (Table 1.10). The range of statistically significant HRs per $10 cost-sharing reduction is 1.027 (18-20 year-olds) to 1.060 (13-14 year-olds). Effect estimates are not significant in the 21-23 or 24-26 year-old age categories.

Socioeconomic correlates of HPV vaccination in the study sample are generally consistent with the literature. As shown Tables 1.3 and 1.8, beneficiaries in ZIP codes with lower median family income and lower high school completion rates are less likely to initiate HPV vaccination, and are also less likely to complete subsequent doses. Beneficiaries in ZIP codes with lower percentages of white residents are more likely to initiate vaccination, but are less likely to complete vaccination conditional on initiating the series; similar differences in HPV vaccination have been reported by individual-level race/ethnicity, both among U.S. adolescents (Reagan-Steiner 2015) and young women (Rahman 2013).

In subgroup analyses, there is evidence of a heterogeneous policy effect with respect to ZIP code-level socioeconomic factors (Table 1.11). In ZIP codes with median 4-person family income ≤200% FPL, the magnitude of the effect estimate for vaccine initiation is roughly double that of the >400% FPL category (5.5% vs. 2.7% increase per $10 cost-sharing reduction; interaction $p<0.001$). Similarly, the effect size for vaccine initiation is over twice as large in the lowest quartile of high school completion than in the highest quartile (6.9% vs. 3.1%; interaction $p<0.001$). Across quartiles of percentage white race/ethnicity, there was evidence of effect modification with respect to two-dose compliance (conditional on vaccine initiation) (interaction $p=0.029$), but not with respect to vaccine initiation (interaction $p=0.132$). The cost-sharing effect

estimate for two-dose compliance is statistically significant in the lowest quartile of percentage white race/ethnicity (2.0% increase per $10 reduction; 95% CI: 0.6-3.4%), as well as in the lowest quartile of education and the second-lowest category of income.

Table 1.10: Effect estimates by attained age: Hazard ratio per $10 cost-sharing reduction

Dependent variable: Years from age-qualification to vaccine initiation

Attained age at time *t*	Person-years at risk	HR (95% CI) in subgroup
9 to 10 years	1,136,060	1.048 (1.03, 1.07)***
11 to 12 years	695,912	1.037 (1.02, 1.05)***
13 to 14 years	414,267	1.060 (1.04, 1.08)***
15 to 17 years	506,503	1.042 (1.03, 1.06)***
18 to 20 years	430,065	1.027 (1.00, 1.05)*
21 to 23 years	301,684	1.000 (0.97, 1.04)
24 to 26 years	176,586	1.014 (0.93, 1.11)

By construction, at every time *t* in the model of years from age-qualification to vaccine initiation, attained age is equal to 9+*t* years. To estimate the cost-sharing effect by age, the primary model specification with plan-level fixed effects is re-run with the addition of interactions between vaccine-related cost-sharing and dummy indicators for each attained age category. A dummy variable for being ≥18 years old and having employee (vs. dependent or spouse) status is also added to the model; this variable is associated with a hazard ratio of 0.85 (95% CI: 0.79, 0.91).
**p<0.05; **p<0.01; ***p<0.001*

Table 1.11: Effect estimates within ZIP code-level socioeconomic subgroups: Hazard ratio per $10 cost-sharing reduction

Subgroup	(i) Dose 1		(ii) Dose 1 to Dose 2		(iii) Dose 2 to Dose 3	
	HR (95% CI) in subgroup	Interaction p-value	HR (95% CI) in subgroup	Interaction p-value	HR (95% CI) in subgroup	Interaction p-value
Median family income (4-person)						
≤200% FPL	1.055 (1.04, 1.07)***	<0.001	1.014 (1.00, 1.03)	0.108	1.007 (0.99, 1.03)	0.128
201-300% FPL	1.062 (1.05, 1.07)***		1.013 (1.00, 1.03)*		1.011 (0.99, 1.03)	
301-400% FPL	1.044 (1.03, 1.05)***		1.008 (1.00, 1.02)		0.997 (0.98, 1.01)	
>400% FPL	1.027 (1.02, 1.04)***		1.006 (0.99, 1.02)		1.007 (0.99, 1.02)	
Quartiles, % adults without high school completion						
1 (highest completion)	1.031 (1.02, 1.04)***	<0.001	1.006 (1.00, 1.02)	0.052	1.007 (0.99, 1.02)	0.252
2	1.041 (1.03, 1.05)***		1.008 (1.00, 1.02)		1.004 (0.99, 1.02)	
3	1.048 (1.04, 1.06)***		1.009 (1.00, 1.02)		1.000 (0.98, 1.02)	
4	1.069 (1.06, 1.08)***		1.017 (1.00, 1.03)*		1.009 (0.99, 1.03)	
Quartiles, % white race						
1 (lowest % white)	1.050 (1.04, 1.06)***	0.132	1.020 (1.01, 1.03)**	0.029	1.001 (0.98, 1.02)	0.331
2	1.040 (1.03, 1.05)***		1.009 (1.00, 1.02)		1.009 (0.99, 1.03)	
3	1.038 (1.03, 1.05)***		1.005 (0.99, 1.02)		1.006 (0.99, 1.02)	
4	1.040 (1.03, 1.05)***		1.007 (1.00, 1.02)		1.004 (0.99, 1.02)	

Dependent variables are: (i) years from age-qualification to dose 1 (initiation); (ii) months from dose 1 to dose 2; and (iii) months from dose 2 to dose 3. For each categorical variable, the primary model specification with plan-level fixed effects is rerun with the addition of interactions between vaccine-related cost-sharing and dummy indicators for each subgroup. Interaction p-values <0.05 indicate significant effect modification, based on Wald chi-square tests. (Asterisks indicate a significant cost-sharing effect within the subgroup.)
**p<0.05; **p<0.01; ***p<0.001*

DISCUSSION

Key findings and policy implications

This paper examines the effect of VBID under U.S. health reform on HPV vaccine uptake and compliance. My analyses take advantage of variation in intensity of the VBID intervention within a large sample of private health plans existing before and after the ACA's enactment. I control for background trends in vaccination and use plan-employer fixed effects to account for unobserved time-constant correlates of vaccine-related cost-sharing and vaccination behavior.

From 2009 to 2013, average cost-sharing per HPV vaccine dose decreased by more than half among study plans. In regression analyses, cost-sharing reductions are associated with statistically significant increases in age-specific vaccine initiation and completion rates in proportion to the magnitude of the reduction from pre-ACA levels. Increases are modest on a per-dollar basis, but could be sizable depending on the plan's pre-ACA cost-sharing scenario. Among girls and women who initiated the vaccine series, improvements in multi-dose compliance with reduced cost-sharing are not statistically significant in the main regression analysis. However, there is evidence of a significant effect on multi-dose compliance in socioeconomic subgroups that have lower background propensity to complete vaccination.

This analysis is among the first to evaluate routine vaccination as a target for VBID. Despite accumulating evidence that VBID programs can moderately improve adherence to chronic disease medications, it was unclear whether to expect similar improvements in HPV vaccine uptake following the ACA's preventive care reforms. In the RAND Health Insurance Experiment, which examined the effect of insurance generosity on health care utilization, cost-sharing decreased overall use of preventive care for children but did not specifically show an effect on immunizations for older children and adolescents. With respect to acute care, cost-sharing decreased care-seeking for children, but did not significantly affect use of highly-effective care for children in non-poor families (Newhouse 1993). Given reports that parents of adolescent girls often have limited knowledge about HPV and the vaccines (Holman 2014), my finding of a significant cost-sharing effect could be the result of parents perceiving HPV vaccines as discretionary or non-essential care. The lack of school mandates for HPV vaccines may reinforce such perceptions and lead HPV immunization rates to be more price-dependent.

The price elasticity of demand in this study (-0.08 for vaccine initiation rates) is comparable to those measured in recent evaluations of VBID programs targeting chronic medication adherence (Chernew 2008; Choudhry 2010; Frank 2012); for example, Frank et al. found an elasticity of -0.06 for adherence to statins in the VBID program of a large regional health plan. Conversely, -0.08 is toward the low end of the elasticity range from studies that broadly examined how beneficiaries' use of health services responds to varying levels of cost-sharing: In a systematic review of such studies, price elasticity of demand for all health services ranged from -0.04 to -0.75, with most estimates clustering around -0.17 (Ringel 2002). As with any VBID intervention, it is possible that limited awareness of the value-based incentives dampened beneficiaries' responsiveness to them. In a 2014 poll, less than half of the population (43%) reported knowing that the ACA eliminated out-of-pocket costs for preventive services (Hamel 2014). The impact of the preventive care reforms on behavior may evolve as more plans lose grandfathered status and the provisions apply more universally to the privately-insured population.

This study contributes to new evidence on the relationship between socioeconomic factors and beneficiaries' sensitivity to cost-sharing changes under VBID. The removal of financial barriers to high-value preventive care should theoretically reduce disparities in the use of these services. However, the literature is mixed on whether and how much heterogeneity in price responsiveness exists across socioeconomic groups (Scitovsky & McCall 1977; Cherkin 1990; Newhouse 1993; Chernew 2008). VBID might even exacerbate inequalities in access if, for example, lower income beneficiaries are less likely to be aware of the benefit variations, as one recent survey found within a VBID program population (Henrikson 2014).

Within study plans, I observe significant variation in HPV vaccination patterns by ZIP code-level socioeconomic factors, as well as variation in price sensitivity by these factors. Beneficiaries in lower categories of ZIP code-level income and educational attainment are less likely to initiate and complete vaccination, and have notably larger increases in vaccine initiation per dollar reduction in cost-sharing. In the most disadvantaged categories of income and education, the price elasticity of vaccine initiation is approximately two times that observed in the least disadvantaged categories. Beneficiaries in the most disadvantaged ZIP code quartile of race/ethnicity are more likely to initiate vaccination compared to the least disadvantaged quartile, but have lower multi-dose compliance conditional on initiation. Improvements in compliance with cost-sharing reductions are statistically significant in this subgroup, despite not being significant in the overall sample. In the absence of household-level data on income, education, and race/ethnicity, it is not possible to distinguish between individual socioeconomic factors and area-level traits in this analysis. Nevertheless, these results are promising and suggest that cost-sharing reductions for the HPV vaccine are likely to help reduce disparities in uptake.

Interestingly, during the post-ACA period, there are large increases in vaccine initiation rates not explained by the vaccine-related cost-sharing reductions since 2009. The increases are seen even among plans with zero or very low baseline cost-sharing levels, which generally had little change in cost-sharing from 2009 through 2013. This trend could be due to unrelated contemporaneous events; for example, ACIP extended the recommendation for routine quadrivalent vaccination to males in mid-2011 (CDC 2011), which may have renewed interest in the vaccine. However, it is also possible that spillover effects of the preventive care provisions contributed to the increase in vaccination rates. In plans without actual cost-sharing changes, the reforms could still have impacted vaccination rates by signaling the value of HPV vaccines to

providers. Announcement of the reforms may have also led patients and providers to assume (correctly or incorrectly) that the HPV vaccine would be fully covered, or alerted more patients to preventive care benefits that their plan already featured before the ACA.

Nevertheless, complementary approaches are probably still needed to substantially raise HPV vaccine coverage among U.S. adolescents and young women. Even under the free vaccination scenario (and even when applying 2013 vaccination rates), cumulative coverage with all three doses still falls far short of federal public health objectives (Healthy People 2020 goal of 80% vaccine completion by ages 13-15). A number of factors hinder HPV vaccine uptake beyond cost, including misinformation about the value of HPV vaccines, discomfort vaccinating young adolescents against a sexually transmitted infection, infrequent contact between adolescent patients and providers, and logistical difficulties of vaccine completion. Additional incremental strategies could focus on improving the quality of provider recommendations (Gilkey 2015), expanding the authority of other provider types (e.g., pharmacists) to deliver HPV vaccines (Brewer 2014), and establishing reminder and recall systems to increase initiation and completion rates (Dunne 2014). The political feasibility of strong interventions (such as HPV immunization requirements for school enrollment) should also be re-evaluated in the wake of health reform, which reduces funding-related barriers to state-mandated vaccinations.

Limitations

This study is subject to several limitations. Plan-level fixed effects greatly reduce the potential for confounding by unobserved plan or group characteristics that independently correlate with HPV vaccine uptake; still, this approach does not exclude the possibility of time-varying confounding. Such bias could occur if reductions in overall generosity tended to

accompany vaccine-related cost-sharing reductions, given the rules defining grandfathered status eligibility. However, the resulting bias would likely be in the opposite direction of the observed policy effect (and small in magnitude, since many plans could have lost grandfathered status voluntarily or due to minor reductions in benefits). Effect sizes show very little change when adding individual-level covariates to a model with plan-employer fixed effects, which helps alleviate concerns about potential changes in the composition of beneficiaries over time.

Although MarketScan includes enrollees from all U.S. states, it is not representative of the entire U.S. private insurance market. The database disproportionately represents large group health plans, which may be more generous on average than small group or individual health plans. Enrollees are also concentrated in more socioeconomically advantaged ZIP codes compared to the general population, and may be less price sensitive than average. Consequently, the results may understate both the extent of the cost-sharing reductions and the per-dollar behavioral effects of the VBID intervention among U.S. health plans.

Empirical cost-sharing estimates are susceptible to random measurement error, which may attenuate the policy effect estimates. However, the large average size of study plans mitigates this concern. The plan-level fixed effects models rely on within-plan changes in cost-sharing, which are likely to have good accuracy in large plans (based on the observed precision of cost-sharing estimates for these plans). Effect sizes only slightly increase in sensitivity analyses that exclude relatively small plans from the sample.

Insured adolescents are eligible to receive free vaccination through the VFC program for any routine vaccines not covered by insurance. As a result, in plans with very high pre-ACA cost-sharing that lost grandfathered status, there could be differentially higher rates of missing HPV vaccination records in years before the ACA-mandated cost-sharing reductions. However,

this concern applies only to plans in the highest pre-ACA cost-sharing stratum (>$50 per dose in 2009), and policy effect estimates show little change when this group is removed from the analyses. Moreover, due to caveats in VFC eligibility for the underinsured, the number of federally-provided vaccinations is likely to be limited even in the >$50 cost-sharing stratum: Insured adolescents are not VFC-eligible for vaccines covered with high coinsurance or only after a deductible is met; and those who are eligible can only receive VFC vaccines through federally-qualified or rural health centers, which have limited capacity and geographic reach (Lindley 2009; Smith 2009).

Conclusions

In years following the ACA's preventive care reforms, plans with larger reductions in vaccine-related cost-sharing tended to show larger increases in HPV vaccine initiation and completion rates. Results suggest that the cost-sharing reductions generated modest increases in vaccine uptake, consistent with prior empirical evaluations of VBID programs targeting prescription drug adherence. Responsiveness to cost-sharing changes is more notable among residents of socioeconomically disadvantaged areas, who have lower background rates of vaccination. Nevertheless, below-target vaccination rates under zero cost-sharing highlight the need for additional interventions to improve HPV vaccine uptake.

Paper 2:

Comparative effectiveness of human papillomavirus vaccination against cervical abnormalities by dose level

INTRODUCTION

Background and significance

Persistent infection with high-risk types of HPV is the cause of essentially all cervical cancer cases (Walboomers 1999; Saslow 2012). HPV is the most common sexually transmitted disease, with most sexually active adults acquiring at least one HPV infection during their lifetime (Weinstock 2004). The quadrivalent and bivalent HPV vaccines protect against the two most carcinogenic strains (HPV-16/18), which account for 70% of cervical cancers worldwide and cause both high-grade cervical lesions (precursors to cancer) and low-grade lesions (de Sanjose 2010). When administered before infection, the vaccines prevent 95-98% of 16/18-related high-grade cervical lesions (Lehtinen 2012; Kjaer 2009), and provide some cross-protection against high-grade lesions related to certain other HPV types (Brown 2009; Schiller 2012; Wheeler 2012). The quadrivalent vaccine also protects against HPV-6/11, which cause over 90% of genital warts (Lacey 2006).

The standard three-dose immunization schedule, however, is expensive and creates logistical challenges for vaccine provision in both low- and high-resource countries (Natunen 2011; Elam-Evans 2014; Fagot 2011; Herweijer 2014). In the U.S., one-third of adolescent girls who initiate vaccination do not complete the series (Elam-Evans 2014). Practical barriers to vaccine completion may also deter providers from recommending and patients from accepting HPV vaccines (Bynum 2014; Ford 2009; Kahn 2008). There are increasing calls to examine the magnitude of protection provided by partial vaccination with one or two doses (Markowitz 2014; President's Cancer Panel 2013; Kreimer 2015). Dose-stratified efficacy data could guide program design by, for example, suggesting the relative priority of strategies to encourage vaccine initiation versus to ensure dosing adherence.

Clinical trials have examined reduced-dose vaccination schedules using immunogenicity or HPV infection as interim outcomes, and generally support the efficacy of partial vaccination (Dobson 2013; Kreimer 2011; Kreimer 2015). Based on immunogenicity findings, a two-dose schedule for young adolescent girls is now licensed in some countries and endorsed by the World Health Organization (WHO 2014). However, further evidence is needed to confirm strong protection against long-term clinical endpoints. Observational data collected since HPV vaccine introduction can help provide such evidence. The U.S. may be an ideal setting for empirical comparisons of varying dose levels given the large number of resident girls with partial HPV vaccination, as well as the low potential for bias from herd immunity.

Specific aims

My paper examines the association between HPV vaccination by dose level and cervical abnormality rates within a large U.S. cohort of privately insured adolescent girls. I compare the incidence of both cytological and histology-confirmed abnormalities between recipients of zero, one, two, or three doses, using inverse probability of treatment weighting (IPTW) to adjust for a broader set of covariates than has previously been possible using conventional regression methods (Crowe 2014; Gertig 2013). Vaccine effect estimates from this study are compared with prior evidence on HPV epidemiology and vaccine efficacy to assess plausibility. Differences in screening initiation rates by HPV vaccination status are also examined.

METHODS

Setting and data source

U.S. guidelines for HPV vaccination and cervical screening

Since 2007, U.S. guidelines have recommended routine three-dose HPV vaccination of 11 to 12 year-old females, with catch-up vaccination of 13 to 26 year-old females who were not previously vaccinated. Vaccination can be started as early as age 9 years (Markowitz 2007). HPV vaccines are primarily targeted to young adolescents because they are most effective when delivered before the onset of sexual activity (Markowitz 2014). To better approximate the potential impact of HPV vaccination in the target age group, my analysis of screening-detected cervical abnormalities focuses specifically on vaccine-eligible adolescent girls who were 9 to 17 years old in 2007.

Cervical cancer screening aims to detect pre-invasive cervical abnormalities, which are asymptomatic (Kumar 2013). Screening guidelines are issued by several different professional organizations[4] in the U.S., and underwent changes during the study timeframe (2007-2013). In 2007, routine Pap-based cytology screening was recommended at either one- or two-year intervals starting at approximately three years after sexual initiation and no later than age 21 years (Saslow 2002; USPSTF 2003; ACOG 2003). As of 2012, all national organizations recommend that women initiate routine Pap screening at age 21 years and continue screening at three-year intervals (Saslow 2012; USPSTF 2012; ACOG 2012).[5]

Claims data

MarketScan Commercial Claims & Encounters (2007-2013) is a de-identified, individual-level database containing health care claims from over 100 employers and insurers, with

[4] Recommendations are issued by the American Cancer Society (ACS), American Society for Colposcopy and Pathology (ASCCP), American Society for Clinical Pathology (ASCP), U.S. Preventive Services Task Force (USPSTF), and American College of Obstetricians and Gynecologists (ACOG).

[5] Guidelines from ACOG were updated in December 2009 to recommend screening starting from age 21 with two-year intervals during ages 21-29 (ACOG 2009); however, guidelines from other organizations were unchanged until 2012.

approximately 33 million enrollees in 2007. Although not nationally representative, the database provides a large convenience sample with enrollees in all U.S. states. Claims are integrated from all outpatient, inpatient, and pharmacy providers that submitted for reimbursement through the enrollee's health plan. With this data, it is possible to track vaccination status and cervical cancer screenings for an enrollee over time. Enrollment files contain basic person-level demographics and plan information, as well as socioeconomic measures based on enrollees' 5-digit ZIP code of residence.

Overview of approach and analytical challenges

Comparator interventions and hypothetical trial design

My analysis seeks to compare cervical abnormality outcomes across four different preventive care regimens: receipt of either zero, one, two, or three HPV vaccine doses followed by new initiation of a Pap screening regimen. In effect, the screening regimen can be viewed as a component of each intervention, albeit one that is common to all four comparators. Each new screening episode for an individual is defined as a Pap cytology test plus any necessary follow-up procedures to confirm or treat atypical findings.

Hypothetically, a randomized controlled trial (RCT) could have been initiated in 2007 to compare the efficacy of these four interventions against cervical abnormalities, assuming no equipoise issues or other practical constraints. In such an RCT, study participants (screening-naïve girls 9-17 years old in 2007) would be randomized to receive zero, one, two or three doses followed by initiation of a cervical cancer screening regimen. Age at the time of randomization is likely to be strongly related to the trajectory of cervical abnormality risk (e.g., due to differences in HPV exposure risk and immune system factors by age); therefore, the randomization could be

conducted separately within blocks defined by starting age in 2007. Age at vaccine initiation would need to be consistent between the intervention groups, which could be accomplished by delivering the first dose at randomization.

Because precancerous cervical lesions are asymptomatic, the schedule of screenings within each age block would also need to be independent of randomization group to prevent differential ascertainment of the outcomes (a type of information bias). The first screening could be scheduled for 2008 or later to provide a sufficient window of time for three-dose vaccination beforehand. With a uniform schedule of screenings across the groups, case-counting for abnormalities could start at the randomization visit as per an intent-to-treat (ITT) approach, with censoring at loss to follow-up.

Observational cohort design

By contrast, in a retrospective cohort study of vaccine effectiveness by dose level, study subjects self-select their number of vaccine doses and the timing thereof. Age at the first screening will vary among girls receiving care in usual practice, especially under pre-2012 guidelines (Saslow 2002). The time interval between consecutive screenings will also vary across (and within) individuals, although to a lesser extent; roughly annual intervals are likely to be common among insured girls and young women during the study timeframe.

As with any observational study, because treatment is not randomly assigned, there is potential for confounding from unobserved factors that affect both treatment choice and outcomes. In addition, because the outcomes in this study are asymptomatic, there is strong potential for outcome ascertainment bias if vaccination status is correlated with the timing of screening. This analytical challenge was recently noted by Herweijer et al. (2015) in a paper that

examined the association between HPV vaccination and screening attendance among Swedish women. After observing significantly higher screening rates among vaccinated women (which was almost entirely explained by socioeconomic factors[6]), the authors noted the following:

> "Our findings […] imply that studies should take increased screening rates in vaccinated women into consideration when assessing the effectiveness against screen-detected cervical cancer precursor lesions […]. If HPV-vaccinated women participate more extensively in screening, the recorded incidence of screen-detected lesions may be paradoxically somewhat increased, compared to the recorded incidence in unvaccinated—and less frequently screened—women. This potential for detection bias may lead to an underestimation of vaccine effectiveness against cervical lesions in population-based studies […]." (Herweijer 2015)

In supplemental analyses, I show that this concern is indeed applicable to the present U.S.-based sample: Among screening-naïve adolescent enrollees, HPV vaccination (≥1 dose) is associated with significantly higher hazard rates of screening initiation relative to no vaccination, although screening rates are similar across the one-, two-, and three-dose levels.

In an effort to minimize the potential for ascertainment bias and mimic other conditions of the hypothetical RCT using retrospective claims, my primary design approach is to: (i) identify screening-naïve female enrollees aged 9-17 years in 2007 who reached their age of first screening during 2008-2013; (ii) assign exposure as the number of vaccine doses received before the first screening; (iii) use an IPTW approach to align the comparator groups in terms of starting age in 2007 and, within each age block, the distribution of other observed characteristics

[6] Among Swedish women invited to attend cervical cancer screening, the hazard ratio of screening attendance associated with vaccination (≥1 dose) versus no vaccination was 1.28 (95% CI: 1.24, 1.32) before covariate adjustment (Herweijer 2015). This hazard ratio decreased to 1.05 (95% CI: 1.02, 1.08) after adjusting for individual-level income and educational attainment, socioeconomic factors that are unavailable in the present U.S.-based analysis.

(including age of first screening and age of vaccine initiation); and (iv) approximately maintain the age/year alignment of dose groups at subsequent Pap visits by censoring individuals upon an 18-month gap without additional screening.

In contrast to an ITT analysis of the hypothetical RCT, in which the time origin for outcome ascertainment is randomization (i.e., vaccine initiation), this observational cohort design instead treats screening initiation as time zero. This time zero assignment resembles that of per-protocol analyses of HPV vaccine trials, which focused on participants who reached the start of the post-vaccination follow-up period without protocol deviations or prior evidence of HPV infection (Kjaer 2009). Note that while this design requires subjects to reach the screening phase of their preventive care regimen, it does not condition on clinical events occurring after the start of screening. Analyses also adjust for age/year of screening initiation so that, at the first screening, the groups are balanced with respect to how long individuals have been screening-naïve/"abnormality-free" thus far.

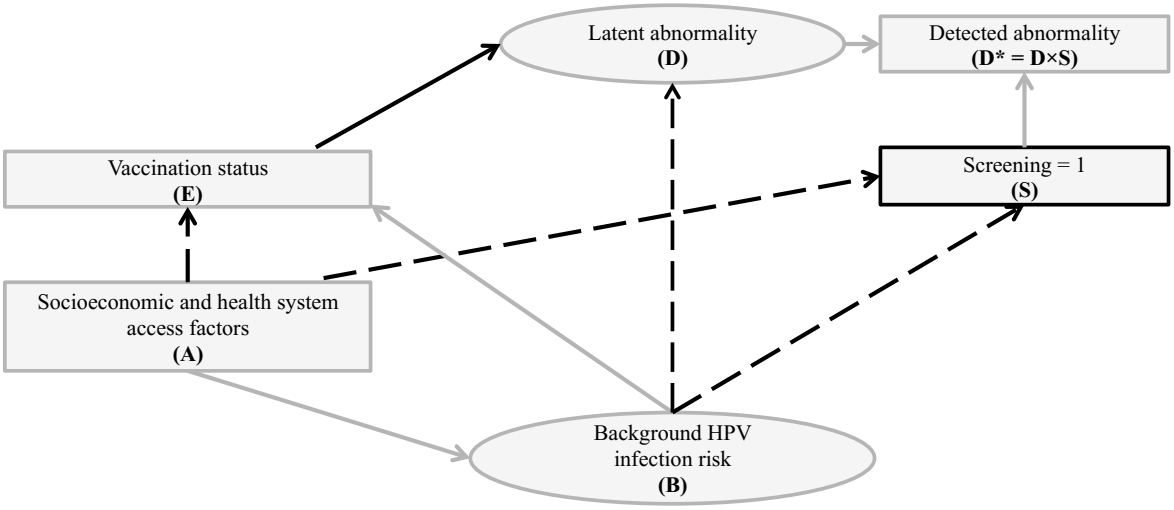
Key limitations

One drawback of this approach is that conditioning sample selection on screening initiation could potentially introduce "M-bias", a form of collider bias that can arise when stratifying on a collider variable (in this case, screening) that is affected by both a risk factor for exposure and a risk factor for disease. Prior evidence suggests that socioeconomic advantage may increase rates of both HPV vaccine exposure (Reagan-Steiner 2015; Rahman 2013) and screening (Herweijer 2015; Leyden 2005). Meanwhile, sexual behavior factors affect the risk of HPV-related cervical atypia, and are also likely to influence screening decisions during the study timeframe due to guidelines that linked the recommended start of screening with age at sexual

initiation (Saslow 2002). As a result, conditioning selection on screening could open a non-causal "back-door" path between HPV vaccination and cervical atypia that biases the estimated effectiveness of vaccination (Greenland 2003).

The simplified schematic in Figure 2.1 illustrates the tradeoff between addressing outcome ascertainment bias versus avoiding collider bias, following the conventions used by Shahar (2009) and Hernán & Cole (2009) to depict information bias in causal diagrams. The diagram shows potential causal pathways linking vaccination and its determinants to the risk of having a cervical abnormality endpoint at a given point in calendar time among girls with the same attained age. The causal effect of interest (represented by the solid black line) is between vaccine exposure status (E) and latent abnormality status (D), an unobserved indicator for the screening test result that would be revealed if the individual were to undergo screening. However, it is only possible to estimate the effect of E on detected abnormality status (D*), which is equal to D if the individual actually does undergo screening (i.e., if S=1) or 0 if the individual remains unscreened (i.e., if S=0).

Figure 2.1: Conceptual diagram of causal pathways linking HPV vaccination status (E) to detected abnormality status (D*) among subjects with the same attained age and birth year



In addition to the direct pathway from E to D* through D (E→D→D*), there are several extraneous pathways linking E to D* through screening (S) that are not of causal interest. Socioeconomic and health system access factors (A) are assumed to influence the likelihood of both vaccination (E) and screening (S). Background risk of HPV infection (B), a latent variable that depends on sexual history and epidemiologic factors, may similarly affect both E and S. As shown, E may also be indirectly connected to B through A, given that the socioeconomic factors affecting vaccination status could also affect background risk of HPV infection.

Adjusting for confounders A and B would block the information biasing pathways from E to D* through S, in addition to blocking the confounding pathways from E to D through A and B; however, full adjustment is not possible using administrative claims data. Therefore, to address information bias, I instead condition sample selection on screening (i.e., S=1), which directly blocks the S→D* path. The cost of this approach is that S becomes a collider variable on the pathway E←A→S←B→D→D*, an M-bias structure represented by the dashed black line. Conditioning on S could thus open up a non-causal, spurious path of association between E and

D* through D. To give a concrete example, it is possible that high socioeconomic status is a cause of both higher vaccination rates and more prompt initiation of cervical screening following sexual initiation. In this case, conditioning on screening initiation could induce a correlation between vaccination and older age of sexual initiation, presumably inducing confounding of the E→D→D* effect by B. If older age at sexual initiation reduces abnormality risk, the resulting bias would make vaccination appear more effective; then again, bias from small differences in age of sexual initiation may be minor (Bahmanyar 2012).

Because the hypothesized collider bias in Figure 2.1 involves variables that are latent (i.e., background risk of HPV) or imperfectly measured (i.e., socioeconomic factors), the magnitude and direction of bias is uncertain. However, I expect that adjusting for screening will result in less bias than failing to adjust for this variable. A paper by Greenland (2003) provides algebraic evidence that M-type collider bias is often slight, and that when a variable is thought to act as both a collider and a confounder, adjustment is likely to remove more bias that it induces. Although screening does not operate as a classical confounder in Figure 2.1, it is a strong determinant of individuals' detected abnormality status and (as reported in the Results section) is significantly correlated with vaccination status despite covariate adjustment. Without adjusting for this path of information bias, analyses comparing vaccination (any dose level ≥1) versus no vaccination are likely to substantially underestimate vaccine effectiveness, as predicted by Herweijer et al. (2015).

Moreover, note that screening acts as a proxy for the confounders socioeconomic status and background HPV risk in Figure 2.1. (This contrasts with usual textbook examples of M-bias, in which the A and B variables are causes of E and S or D and S, but are not also confounders

that cause both E and D.) Consequently, adjusting for screening may reduce more confounding than it induces, in addition to the advantage of this approach in preventing information bias.

To reduce the potential for both selection-induced and classical confounding, I control for a variety of observed demographic, socioeconomic, and utilization-based covariates likely to be correlated with vaccination status, cervical abnormality risk, and/or screening participation. In a sensitivity analysis, I compare cervical abnormality rates across the one-, two-, and three-dose levels using an alternative cohort design that does not condition sample selection on screening initiation. Because residual confounding remains a concern, particularly for comparisons with the zero-dose group, my paper also explores the plausibility of vaccine effect estimates using prior evidence on vaccine efficacy and disease epidemiology.

Sample selection

Within the MarketScan database, the retrospective cohort includes females born during 1989-1997 who were continuously enrolled for at least 2007 and 2008. Per the hypothetical RCT design, I select enrollees with one or more Pap screening claims, with the first occurring in 2008 or later. To increase confidence that subjects were screening naïve before their first observed Pap test in the database, I exclude enrollees with prior evidence of screening, including those with HPV DNA or histology (i.e., follow-up diagnostic) procedure claims before the first Pap test date. Those who began screening before age 11 years are also excluded due to small numbers.

To ensure a similar schedule of screenings across vaccine exposure groups, the outcome ascertainment period begins with the first Pap screening and continues until the first screening with the abnormality of interest, or the last screening before right censoring. I account for interval censoring by restricting the main analysis to consecutive Pap visits occurring at

approximately annual intervals (9-18 months), which reduces variability in the spacing of screenings beyond the first visit. (Pap tests occurring less than 9 months from the previous included Pap screening are regarded as repeat cytology rather than the start of a new screening episode; thus, any diagnoses associated with these repeat cytology tests are attributed to the preceding Pap screening.) Individuals are right censored at the earliest of: December 31, 2013, disenrollment from the MarketScan database, or an 18-month gap without Pap screening.

Variable measurement

Exposure definitions

Vaccination records are extracted from both medical and pharmacy claims (Appendix B.1). I define vaccine exposure by the number of doses received before the first Pap test date, counting all vaccine claims that complied with minimum acceptable intervals between doses (24 days between doses one and two, 80 days between doses two and three) (CDC 2012). Vaccine exposure assignment is static rather than time-varying due to the expected influence of screening results on subsequent vaccination decisions, as well as the higher likelihood of pre-existing HPV infections by the typical age of screening.

Outcome measures

To identify abnormalities detected within each screening episode, I use diagnosis codes associated with the initial Pap screening (i.e., the visit that marks the beginning of a new screening episode) and any follow-up procedures within 9 months of that visit (Appendix B.1). Follow-up could include repeat cytology, or histology procedures (e.g., colposcopy, biopsy, cervical excision) to diagnose or treat abnormal cytology findings.

I categorize abnormalities by cytology result (according to the 2001 Bethesda System) and by follow-up histology grade (Solomon 2002). Cytological abnormalities include: high-grade squamous intraepithelial lesion (HSIL), low-grade SIL (LSIL), atypical glandular cells (AGC), atypical squamous cells of undetermined significance (ASC-US), and ASC, cannot rule out HSIL (ASC-H). In the absence of abnormal cytology diagnoses, an unknown cytological abnormality is flagged if the Pap test prompted histology follow-up. I classify histology-confirmed cervical intraepithelial neoplasia (CIN) as either high-grade (CIN2 or CIN3+) or low-grade (CIN1 or unspecified dysplasia). If more than one histology grade is recorded within the same screening episode, the higher-severity grade is used.

Primary study endpoints include any cytological abnormality (a composite measure) and high-grade CIN. As secondary endpoints, each cytological abnormality type and histology grade is examined individually in order to assess variation in effect sizes by endpoint severity.

Covariates

I adjust for birth year, age at first screening, age at vaccine initiation, and years since vaccine initiation as of the first screening, plus additional demographic and socioeconomic covariates: census division of residence; data contributor type (insurer or employer); health plan type; outpatient utilization in 2007 (≥1 visit) as a measure of health system access; and ZIP-code-level categories of percentage white race/ethnicity, percentage without high school completion, and median family income. I also include two claims-based proxies for potential HPV infection risk (history of testing and history of diagnosis for chlamydia or gonorrhea at or before the first screening visit), based on their empirical correlations with cytological abnormality risk (Appendix B.2).

Statistical analysis

Primary analysis

I conduct discrete-time survival analyses using pooled logistic regression models to estimate the incidence rate of cervical abnormalities by vaccine dose level. The models are fitted using IPTW to reweight subjects such that the distributions of observed covariates in each exposure group closely match that of the entire study cohort (Austin 2011). IPTW is a propensity score-based adjustment method that offers a more practical alternative to propensity score matching for analyses featuring multi-level (>2) treatment groups. Compared to conventional regression analysis, IPTW (like other propensity score methods) generally allow for more flexible covariate selection when treatment is common and binary outcomes are rare (Braitman & Rosenbaum 2002), as in the present study. IPTW also enables me to simultaneously compare outcomes across all dose levels (including the unvaccinated group), while controlling for attributes of the HPV vaccination regimen that are only applicable to recipients of ≥1 dose (e.g., age at first dose).

I stratify the propensity score estimation by birth year to adjust for age in 2007 and the age-specific distributions of all other covariates. Within each birth year stratum, I use a three-level nested logit model to estimate the probabilities of having received zero, one, two, or three doses before screening. The first logistic regression estimates the probability of having received ≥1 (versus 0) doses before screening initiation, given the calendar year of the first Pap, demographic and socioeconomic factors, and claims-based risk proxies. The second model estimates the probability of having received ≥2 doses among women with ≥1 dose, given the above covariates plus variables for age at first dose and the time interval between the first dose

and the first Pap test. The third model includes the same covariates as the second model, and estimates the probability of vaccine completion among those with ≥2 doses. Each individual is assigned a weight equal to the inverse of her estimated propensity to receive her actual number of doses. To avoid assigning extreme weights to individuals, weights are truncated at the 99th percentile in each group.

Pooled logistic regressions (with person-screening as the unit of analysis) are fitted with the weights to estimate the relationship between dose level and incidence of abnormalities (Hernán 2000). For each abnormality endpoint, the regression models the hazard probability of detection within a screening episode, given that the abnormality had not been detected by the previous screening. Independent variables include indicators for dose level and episode number (the individual's first, second, or nth screening). This specification allows the underlying hazard probability of abnormalities to vary over time while assuming proportional hazard probabilities between the exposure groups at any specific visit number, analogous to a Cox proportional hazards model for continuous-time survival outcomes. To estimate the adjusted incidence rate of an abnormality outcome by dose level, I average the hazard probability associated with each dose level over the distribution of visit numbers in the overall cohort. I use the hazard probability odds ratio comparing different dose levels to approximate the relative risk and summarize the results as point estimates of the incidence rate (events per 100 person-screenings), relative risks for all pairwise comparisons, and associated 95% confidence intervals (CIs). All models use robust variance estimation.

Crude incidence rates of abnormalities (equal to the number of events divided by person-visits at risk) are similarly analyzed using un-weighted pooled logistic regressions that only include indicators for dose level.

Assessing plausibility and sensitivity of results

Comparing effect estimates with prior epidemiologic data

To examine the plausibility of my effect estimates, I compare the risk reductions for full vaccination (versus no vaccination) with prior evidence from epidemiologic studies (Insinga 2008; Hariri 2012) and multinational RCTs of the vaccines among 15-26 year-old women (Lehtinen 2012; Kjaer 2009).

Specifically, I first compare these effect estimates with the reported proportions of U.S. CIN cases attributed to vaccine-targeted HPV types. Because these attributable proportions were estimated before HPV vaccines became widely available, they provide a theoretical maximum for the percent of CIN cases that are vaccine-preventable (assuming no cross-protection against other HPV-related CIN and no replacement of HPV types in the population). I interpret these comparisons based on reported vaccine efficacy against vaccine type-related CIN1, CIN2+, and CIN3+.

Additionally, I compare my effect estimates with the reported efficacy of the quadrivalent and bivalent vaccines against *all-cause* CIN1, CIN2+, and CIN3+ (i.e., CIN cases caused by any HPV type). These efficacy results can be directly compared with effect sizes from the present study, with the caveat that efficacy against all-cause CIN partly depends on the proportions attributable to vaccine HPV types within the mix of countries that contributed trial participants.

Testing different exposure or outcome ascertainment windows

I conduct a series of sensitivity analyses to assess robustness. First, I restrict outcome ascertainment to subjects' first screening episode only. The main rationale for this test is to

ensure that the risk reductions observed with partial vaccination in the main analyses are not driven by the subset of girls who completed vaccination during follow-up (3.2%, 5.7%, and 9.1% of the zero-, one-, and two-dose groups, respectively). This sensitivity analysis also addresses the possibility of informative censoring in the main analysis, i.e., if subjects who continued annual Pap screening beyond the first Pap exam are not representative of subjects with the same baseline covariate values who were right-censored.

Second, I expand follow-up to include screening episodes that occurred at any interval of at least 9 months. The main analysis focuses on screenings that occurred approximately annually (9-18 months), which could potentially overstate vaccine impact if a short screening interval is more likely to detect transient abnormalities that otherwise would have resolved before detection.

Third, I apply buffer periods for vaccine exposure assignment to reflect the uncertain latent period between HPV infection and abnormality development. In this set of analyses, dose level is assigned according to the number of doses received prior to a buffer period (30, 90 or 180 days) before the first Pap cytology visit. The rationale is to omit doses received shortly before the start of outcome ascertainment, with the assumption that any detected abnormalities are likely due to HPV infections that pre-existed those vaccine doses. For each alternative exposure definition, I re-estimate subjects' inverse probability of treatment weights before fitting the weighted logistic regression models.

Examining potential screening-related selection bias

The main analyses adjust for the timing, spacing, and number of screenings to prevent bias from differential outcome ascertainment; however, as discussed earlier, the decision to

condition sample selection on screening (a potential collider) may create selection bias. I therefore conduct additional analyses to gauge the potential for selection bias due underlying differences in screening participation across exposure groups.

First, I evaluate the association between HPV vaccination status and time to cervical screening initiation in the broader sample of age-eligible female enrollees (9-17 years old in 2007) who were enrolled at least for the 2007 and 2008 plan years and were screening-naïve as of January 1, 2008. I fit a Cox proportional hazards model of time to first Pap screening (measured in months from January 2008). Follow-up continues until the month of first screening or censoring at disenrollment or December 2013. Given the importance of starting age on time to first screening, I stratify the baseline hazard function by age in 2007 (i.e., by birth year). Independent variables include time-varying indicators for vaccination status (corresponding to the number of doses received up to and including each month of follow-up), and socioeconomic and demographic covariates.

Second, I conduct a sensitivity analysis of cervical abnormality rates with full versus partial vaccination using an alternative cohort study design, which does not require screening initiation for cohort entry. This analysis includes age-eligible girls who initiated HPV vaccination prior to screening initiation or disenrollment, whichever occurred first. Follow-up starts at vaccine initiation, with the assumption that time prior to screening was time without the abnormality. (As described in the Results section, timing of the first Pap exam is non-differential across the one-, two-, and three-dose levels; ascertainment bias therefore is not a major concern for the comparison of dose levels ≥1.) I fit Cox proportional hazard models of months from vaccine initiation until abnormality detection or censoring, with time-varying indicators for doses

received up to and including each month of follow-up.[7] The model is stratified by each unique combination of age and calendar year at vaccine initiation. Demographic and socioeconomic factors are included as covariates.

Subgroup analyses

Because the privately insured cohort that I examine is concentrated in more socioeconomically advantaged ZIP codes compared to the general population, I test for effect modification with respect to outpatient utilization, ZIP code-level socioeconomic measures, and sexually transmitted infection testing. In subgroup analyses of the two primary study endpoints, I estimate vaccine effectiveness by dose level within each strata by adding the relevant main effect and its interactions with treatment to the weighted pooled logistic regression.

RESULTS

Cohort characteristics

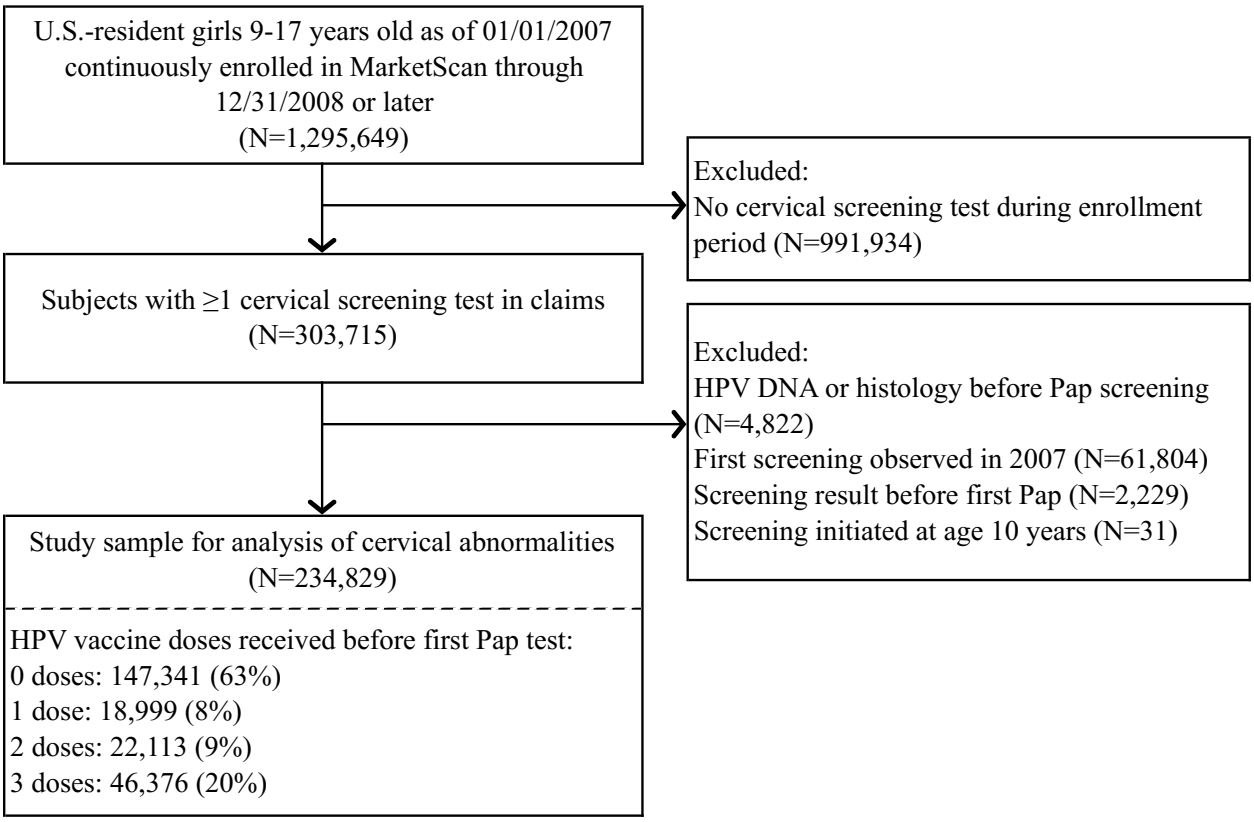
The screening cohort includes 234,829 adolescent girls who received zero (63%, N=147,341), one (8%, N=18,999), two (9%, N=22,113), or three (20%, N=46,376) HPV vaccine doses before their first cervical screening (Figure 2.2). Over 99% of those vaccinated received the quadrivalent vaccine.

On average, study subjects initiated screened at 17.8 years old and (among those with ≥1 dose) initiated vaccination at 15.6 years old (Table 2.1). Compared to fully vaccinated girls, partially vaccinated girls were generally older at vaccine initiation. Girls with fewer or no

[7] Due to endogeneity concerns, and because the interventions of interest consist of vaccine doses administered prior to screening, vaccine doses received at or after the first Pap screening visit do not count towards an individuals' vaccination status in the Cox regression analysis. To account for the portion of individuals who receive additional doses post-screening initiation, I fit another set of Cox regression models in which individuals are right-censored beyond their first screening episode.

vaccine doses also had lower outpatient utilization and tended to reside in areas with lower median family income and high school completion rates. Prevalence of testing and diagnoses for chlamydia or gonorrhea were highest among one- and two-dose recipients. In the reweighted sample, observed characteristics were well matched across the four groups (Table 2.1).

Figure 2.2: Selection of screening cohort to examine cervical abnormality rates by HPV vaccine dose level



Vaccine claims count as separate dosages if time since the previous dose complied with minimum intervals (24 days between doses one to two, 80 days between doses two to three, i.e., 4 weeks and 12 weeks minus an extra 4-day grace period).

Table 2.1: Cohort characteristics before and after inverse probability of treatment weighting

Covariate	Before weighting				After weighting[a]			
	0 dose	1 dose	2 dose	3 dose	0 dose	1 dose	2 dose	3 dose
N	147,341	18,999	22,113	46,376	147,341	18,999	22,113	46,376
Mean age in 2007 (SD)	15.1 (1.6)	14.9 (1.7)	14.9 (1.7)	14.8 (1.7)	15.0 (1.7)	15.0 (1.7)	15.0 (1.7)	15.0 (1.7)
Mean age at first Pap (SD)	17.7 (1.9)	17.6 (1.9)	17.7 (1.9)	18.0 (1.9)	17.7 (1.9)	17.8 (1.9)	17.8 (1.9)	17.8 (1.9)
Year of first Pap, %								
2008	33.4	30.1	27.5	16.7	29.3	27.9	28.2	27.5
2009	26.2	25.3	25.1	23.7	25.6	25.4	25.5	25.2
2010	13.6	13.9	13.9	16.6	14.2	14.1	14.5	14.6
2011	10.4	11.1	12.0	15.0	11.5	11.8	11.8	12.0
2012	9.3	10.6	11.5	14.8	10.7	11.3	10.9	11.3
2013	7.1	8.9	10.0	13.3	8.7	9.4	9.1	9.3
Census division, %								
East North Central	21.3	18.2	19.3	22.1	21.0	20.8	20.9	20.6
East South Central	8.1	6.5	6.6	7.2	7.7	7.6	7.6	7.7
Middle Atlantic	4.7	7.1	8.4	9.8	6.3	6.4	6.4	6.5
Mountain	4.5	4.4	4.2	3.6	4.3	4.5	4.4	4.3
New England	1.8	3.8	5.4	6.5	2.9	3.2	3.3	3.3
Pacific	11.4	18.1	15.1	10.6	12.3	12.7	12.4	12.8
South Atlantic	21.8	20.7	21.0	20.2	21.3	20.9	21.3	21.2
West North Central	5.6	5.5	6.0	7.6	6.0	5.9	5.9	6.0
West South Central	20.9	15.8	14.1	12.5	18.2	18.1	17.8	17.6
Data contributor type (insurer vs. employer)[b], %	39.5	34.5	33.6	30.5	36.9	36.3	36.3	35.7
Plan type, %								
Health maintenance organization	18.1	23.4	22.1	21.2	19.6	19.4	19.6	19.8
Point-of-service (non-capitated)	11.2	11.9	13.0	13.9	11.8	11.8	12.0	12.0
Preferred provider organization	62.9	57.2	57.0	57.1	60.8	61.0	60.6	60.4
Other plan types[c]	7.8	7.6	7.9	7.8	7.8	7.8	7.9	7.8
Any outpatient service use in 2007, %	82.4	90.9	94.5	96.0	86.9	87.6	88.8	90.3
Socioeconomic measures by ZIP code[d], %								
Median family income (4-person)								
≤200% FPL	10.9	8.6	7.1	5.3	9.3	9.1	8.9	8.6
201-300% FPL	29.2	24.0	21.0	18.6	26.0	26.0	25.7	25.6
301-400% FPL	29.0	27.1	26.4	26.1	28.1	28.4	28.2	28.0
>400% FPL	28.8	38.0	43.2	48.0	34.5	34.4	35.0	35.7
Quartiles, % adults without high school completion								

Table 2.1 (Continued)

1 (highest % completion)	26.2	33.2	38.6	44.0	31.2	31.5	31.7	32.1
2	25.2	25.2	25.2	25.1	25.3	25.4	25.4	25.4
3	26.1	22.4	20.0	18.0	23.7	23.6	23.6	23.7
4	20.3	16.9	13.9	10.9	17.6	17.5	17.1	16.8
Quartiles, % white race								
1 (lowest % white)	20.3	21.1	17.5	14.2	19.0	19.0	18.6	18.3
2	25.6	26.6	25.9	25.0	25.7	25.6	25.7	25.9
3	27.2	28.2	30.0	32.0	28.3	28.6	28.4	28.6
4	24.7	21.8	24.1	26.9	24.8	24.7	25.2	25.1
Missing socioeconomic measures	2.1	2.3	2.4	2.0	2.1	2.1	2.1	2.0
Claim(s) at or before first Pap[e], %								
Chlamydia or gonorrhea testing	55.0	62.2	60.1	58.9	56.7	56.8	56.9	56.9
Chlamydia or gonorrhea diagnosis	0.6	0.7	0.7	0.5	0.6	0.6	0.6	0.6
Mean age of vaccine initiation (SD)	-	16.0 (1.7)	15.6 (1.7)	15.3 (1.7)	-	15.7 (1.7)	15.7 (1.7)	15.6 (1.7)
Time since vaccine initiation at first Pap, %								
<1 year	-	42.8	27.0	9.7	-	26.2	25.9	23.4
1-2 years	-	24.3	28.2	29.4	-	30.6	30.6	31.7
>2 years	-	32.9	44.7	60.9	-	43.2	43.4	44.9

[a] Subjects are weighted by the inverse of their estimated propensity to be in their observed exposure group.
[b] Type of MarketScan data contributor (insurer vs. employer) is a proxy for smaller versus larger employer size, respectively.
[c] Other plan types are combined above but are adjusted for individually: comprehensive, exclusive provider organization, point-of-service (with capitation), consumer-driven health plan, and other/unknown.
[d] Categorical socioeconomic measures at the 5-digit ZIP code level are based on the American Community Survey 5-year estimates (2008-2012). Median 4-person family income is compared to federal poverty level (FPL) for a 4-person household in 2012 ($23,050).
[e] Chlamydia and gonorrhea are sexually-transmitted infections for which routine testing is recommended among sexually-active girls and young women (Workowski 2015).

3.2 Incidence of cervical abnormalities

The numbers of events and annual person-visit contributions by dose level are presented in Table 2.2 for the primary study endpoints, along with crude incidence rates. As shown, crude rates of any cytological abnormality fall as the number of doses increases (Table 2.2).

The adjusted rates of any cytological abnormality (per 100 annual person-screenings) are similar to unadjusted rates, ranging from 11.60 with zero doses to 9.05 with three doses (Table 2.3). Relative risks (95% CI) of any cytological abnormality versus zero doses are 0.91 (0.87-0.95), 0.84 (0.81-0.88), and 0.76 (0.73-0.79) for one, two, and three doses, respectively. For histologically confirmed high-grade CIN, relative risks (95% CI) versus zero doses are 0.61 (0.50-0.75), 0.68 (0.57-0.82) and 0.47 (0.40-0.55) with one, two, and three doses.

For most abnormality endpoints, I observe significant risk reductions versus no vaccination starting at the one-dose level. The risk reductions associated with vaccination are largest for the highest-severity endpoint, CIN3+ (e.g., relative risk for three versus zero doses: 0.33; 95% CI: 0.23-0.46). Effect sizes are considerably smaller for low-severity endpoints such as LSIL, ASC-US, and low-grade CIN.

Point estimates of the incidence rate tend to decrease with increasing dose level for most abnormality types. However, the relative risks comparing one-, two-, and three-dose vaccination are significant for only a subset of endpoints. Compared to two doses, full vaccination is associated with significantly lower rates of both primary study endpoints (both $p<0.05$). Compared to one dose, two doses is associated with significantly lower rates of any cytological abnormality ($p<0.05$).

Table 2.2: Unadjusted number of events and person-visits by HPV vaccine dose level: Any cervical abnormality and histology-confirmed high-grade CIN

Endpoint	Persons, No.	Person-time at risk (Annual screenings, No.)	Events, No.	Unadjusted incidence rate (95% CI)[a] (Events per 100 annual person-screenings)
Any cytological abnormality				
0 dose (unvaccinated)	147,341	202,838	23,876	11.77(11.63, 11.91)
1 dose	18,999	26,233	2,793	10.65(10.28, 11.02)
2 dose	22,113	31,822	3,110	9.77(9.45, 10.10)
3 dose	46,376	67,841	6,100	8.99(8.78, 9.21)
High-grade CIN (CIN2+)[b]				
0 dose (unvaccinated)	147,341	215,256	1,605	0.75(0.71, 0.78)
1 dose	18,999	27,778	124	0.45(0.37, 0.53)
2 dose	22,113	33,628	157	0.47(0.40, 0.55)
3 dose	46,376	71,364	232	0.33(0.29, 0.37)

[a] Results are from unweighted pooled logistic regression models with person-screening as the unit of analysis.
[b] High-grade cervical intraepithelial neoplasia (CIN) includes CIN2, CIN3, adenocarcinoma in situ, and rare instances of cervical malignancy.

Table 2.3: Adjusted rates of cytological and histology-confirmed abnormalities by HPV vaccine dose level

Endpoint	Adjusted incidence (per 100 annual person-screenings), by dose level[a]				Adjusted relative risk (95% CI)					
	0	1	2	3	1 vs. 0	2 vs. 0	3 vs. 0	3 vs. 2	3 vs. 1	2 vs. 1
By cytology result[b]										
Any abnormality	11.60	10.64	9.93	9.05	0.91 (0.87, 0.95) ***	0.84 (0.81, 0.88) ***	0.76 (0.73, 0.79) ***	0.90 (0.86, 0.95) ***	0.84 (0.79, 0.88) ***	0.93 (0.87, 0.98) *
HSIL	0.43	0.33	0.26	0.24	0.77 (0.60, 0.98) *	0.59 (0.46, 0.75) ***	0.57 (0.47, 0.69) ***	0.96 (0.72, 1.29)	0.74 (0.55, 0.99) *	0.77 (0.55, 1.07)
LSIL	4.81	4.46	3.83	3.60	0.92 (0.86, 0.99) *	0.79 (0.74, 0.84) ***	0.74 (0.70, 0.78) ***	0.94 (0.87, 1.02)	0.80 (0.74, 0.87) ***	0.85 (0.78, 0.93) ***
AGC	2.92	2.71	2.53	2.13	0.93 (0.85, 1.01)	0.86 (0.80, 0.94) ***	0.72 (0.68, 0.78) ***	0.84 (0.76, 0.93) ***	0.78 (0.70, 0.87) ***	0.93 (0.83, 1.04)
ASC-H	0.75	0.67	0.56	0.59	0.89 (0.75, 1.06)	0.74 (0.63, 0.87) ***	0.79 (0.69, 0.90) ***	1.06 (0.87, 1.30)	0.88 (0.72, 1.09)	0.83 (0.66, 1.04)
ASC-US	5.57	5.06	4.78	4.65	0.90 (0.85, 0.96) **	0.85 (0.80, 0.90) ***	0.83 (0.79, 0.87) ***	0.97 (0.90, 1.04)	0.92 (0.85, 0.99) *	0.94 (0.87, 1.02)
Unknown type	0.93	0.83	0.84	0.70	0.90 (0.77, 1.04)	0.90 (0.78, 1.03)	0.75 (0.67, 0.84) ***	0.83 (0.71, 0.98) *	0.84 (0.70, 1.00)	1.01 (0.83, 1.22)
Any histology follow-up	6.38	5.68	5.24	4.81	0.88 (0.83, 0.94) ***	0.81 (0.77, 0.86) ***	0.74 (0.71, 0.78) ***	0.91 (0.85, 0.98) *	0.84 (0.78, 0.90) ***	0.92 (0.85, 0.99) *
By histology grade[c]										
High-grade CIN (CIN2+)	0.72	0.44	0.49	0.34	0.61 (0.50, 0.75) ***	0.68 (0.57, 0.82) ***	0.47 (0.40, 0.55) ***	0.69 (0.54, 0.87) **	0.77 (0.60, 0.99) *	1.12 (0.86, 1.46)
CIN3+	0.21	0.11	0.12	0.07	0.50 (0.33, 0.76) **	0.57 (0.40, 0.82) **	0.33 (0.23, 0.46) ***	0.57 (0.35, 0.92) *	0.65 (0.39, 1.09)	1.14 (0.67, 1.95)
CIN2	0.52	0.34	0.36	0.27	0.65 (0.52, 0.82) ***	0.70 (0.57, 0.86) ***	0.53 (0.44, 0.63) ***	0.75 (0.57, 0.97) *	0.81 (0.61, 1.07)	1.08 (0.80, 1.45)
Low-grade CIN	2.75	2.50	2.37	2.22	0.91 (0.83, 0.99) *	0.86 (0.79, 0.93) ***	0.80 (0.75, 0.86) ***	0.93 (0.85, 1.03)	0.88 (0.80, 0.98) *	0.95 (0.84, 1.06)
CIN1	2.14	1.95	1.82	1.74	0.91 (0.82, 1.00)	0.85 (0.77, 0.93) ***	0.81 (0.75, 0.87) ***	0.96 (0.85, 1.07)	0.89 (0.79, 1.01)	0.93 (0.82, 1.06)
Unspecified	0.65	0.60	0.57	0.52	0.93 (0.78, 1.11)	0.88 (0.75, 1.03)	0.80 (0.69, 0.92) **	0.91 (0.74, 1.11)	0.86 (0.69, 1.06)	0.95 (0.76, 1.19)

[a] A subject can only contribute up to one event per endpoint category, but may appear in more than one category. Results are from weighted pooled logistic regression models with person-screening as the unit of analysis. Robust variance estimation is used.

[b] Cytological abnormalities are classified according to the 2001 Bethesda System and include: high-grade squamous intraepithelial lesion (HSIL), low-grade SIL (LSIL), atypical glandular cells (AGC), atypical squamous cells of undetermined significance (ASC-US), and ASC, cannot rule out HSIL (ASC-H). Unknown abnormality is flagged if the Pap test had no recorded cytological diagnosis but prompted histology follow-up within 9 months.

[c] Histology-confirmed cervical intraepithelial neoplasia (CIN) is identified by diagnosis codes associated with follow-up histology procedure(s). If more than one histology grade is recorded within the same screening episode, the higher-severity grade was used. CIN3+ includes CIN3, adenocarcinoma in situ, and rare instances of cervical malignancy.

*$p<0.05$; **$p<0.01$; ***$p<0.001$

Sensitivity analyses

Results are similar in sensitivity analyses that vary the exposure or outcome ascertainment window (Table 2.4). When focusing on subjects' first screening episode only, one dose and two doses continue to be associated with significantly lower risks of any cytological abnormality and high-grade CIN versus zero doses, despite the reduced statistical power in this analysis. This result offers assurance that the estimated benefit of partial vaccination in the main analysis is not driven by subjects who completed vaccination at or after their first Pap visit. Additionally, the study findings are not sensitive to the inclusion of cervical screenings that occurred at longer than annual intervals. Findings are also comparable when using varying buffer periods for vaccine exposure assignment, although the relative risks comparing full versus partial vaccination regimens are not statistically significant across all buffer periods with respect to high-grade CIN.

I find no evidence of effect modification across socioeconomic and risk proxy subgroups with respect to any cytological abnormality (Table 2.5) or high-grade CIN (Table 2.6).

Appendix B.3 reports parameter estimates from the Cox regression model of time to first cervical screening among age-eligible adolescent girls (N=1,226,763). Based on this model, Table 2.7 provides risk-adjusted cumulative incidences of screening initiation by dose number. The hazard of screening initiation is approximately 25% lower among unvaccinated versus vaccinated (≥1 dose) individuals, which translates to a cumulative incidence of 31% versus 37% through 2013, respectively. Screening rates are otherwise non-differential by dose. Apart from birth year, the single strongest predictor of screening initiation in the model is outpatient service utilization (hazard ratio: 1.57; 95% CI: 1.55-1.59).

Across the ≥1 dose levels, the study conclusions are robust when using an alternative study design that does not condition sample selection on screening initiation. In the Cox models of time from vaccine initiation to abnormality detection (N=411,312), hazard ratios for full versus partial vaccination are comparable to relative risks from the main analysis with respect to both primary endpoints (Table 2.8a). Similarly, when I modify the Cox models to censor individuals beyond their first cervical screening episode (Table 2.8b), hazard ratios for full versus partial vaccination are consistent with relative risks from the logistic regression analysis focusing on screened individuals' first screening episode only (sensitivity analysis 1 in Table 2.4).

Table 2.4: Sensitivity analyses using alternative vaccine exposure or outcome ascertainment window

Endpoint / Version of analysis	Adjusted prevalence (%) or incidence (per 100 person-screenings), by dose level[a]				Adjusted relative risk (95% CI)					
	0	1	2	3	1 vs. 0	2 vs. 0	3 vs. 0	3 vs. 2	3 vs. 1	2 vs. 1
Any abnormality										
Main analysis	11.60	10.64	9.93	9.05	0.91 (0.87, 0.95)***	0.84 (0.81, 0.88)***	0.76 (0.73, 0.79)***	0.90 (0.86, 0.95)***	0.84 (0.79, 0.88)***	0.93 (0.87, 0.98)*
SA 1: First screening[b]	11.41	10.53	9.55	8.81	0.91 (0.86, 0.97)**	0.82 (0.78, 0.86)***	0.75 (0.72, 0.78)***	0.92 (0.86, 0.98)**	0.82 (0.77, 0.88)***	0.90 (0.83, 0.97)**
SA 2: Screenings of all intervals[c]	11.78	10.91	10.18	9.38	0.92 (0.88, 0.96)***	0.85 (0.82, 0.88)***	0.78 (0.75, 0.80)***	0.91 (0.87, 0.96)***	0.85 (0.80, 0.89)***	0.93 (0.88, 0.98)**
SA 3: Buffer for dose assignment[d]										
30-day buffer period	11.59	10.60	9.90	9.06	0.90 (0.86, 0.95)***	0.84 (0.80, 0.88)***	0.76 (0.73, 0.79)***	0.91 (0.86, 0.96)***	0.84 (0.79, 0.89)***	0.93 (0.87, 0.99)*
90-day buffer period	11.58	10.52	9.88	9.06	0.90 (0.86, 0.94)***	0.84 (0.80, 0.87)***	0.76 (0.73, 0.79)***	0.91 (0.86, 0.96)***	0.85 (0.80, 0.90)***	0.93 (0.88, 0.99)*
180-day buffer period	11.54	10.37	9.85	8.93	0.89 (0.84, 0.93)***	0.84 (0.80, 0.87)***	0.75 (0.72, 0.78)***	0.90 (0.85, 0.95)***	0.85 (0.80, 0.90)***	0.94 (0.89, 1.01)
High-grade CIN (CIN2+)										
Main analysis	0.72	0.44	0.49	0.34	0.61 (0.50, 0.75)***	0.68 (0.57, 0.82)***	0.47 (0.40, 0.55)***	0.69 (0.54, 0.87)**	0.77 (0.60, 0.99)*	1.12 (0.86, 1.46)
SA 1: First screening[b]	0.61	0.44	0.47	0.30	0.72 (0.56, 0.92)*	0.76 (0.61, 0.95)*	0.50 (0.40, 0.62)***	0.65 (0.48, 0.88)**	0.69 (0.50, 0.95)*	1.06 (0.77, 1.47)
SA 2: Screenings of all intervals[c]	0.73	0.46	0.51	0.37	0.63 (0.52, 0.76)***	0.70 (0.59, 0.82)***	0.50 (0.43, 0.58)***	0.71 (0.58, 0.88)**	0.79 (0.63, 0.99)*	1.11 (0.88, 1.41)
SA 3: Buffer for dose assignment[d]										
30-day buffer period	0.71	0.43	0.48	0.35	0.60 (0.49, 0.74)***	0.68 (0.56, 0.81)***	0.49 (0.41, 0.58)***	0.72 (0.57, 0.92)**	0.81 (0.63, 1.05)	1.13 (0.87, 1.47)
90-day buffer period	0.71	0.45	0.43	0.34	0.64 (0.52, 0.78)***	0.60 (0.50, 0.73)***	0.48 (0.40, 0.58)***	0.80 (0.62, 1.04)	0.76 (0.58, 0.99)*	0.94 (0.72, 1.23)
180-day buffer period	0.71	0.42	0.42	0.32	0.59 (0.47, 0.72)***	0.60 (0.49, 0.72)***	0.45 (0.38, 0.54)***	0.76 (0.59, 0.98)*	0.77 (0.59, 1.01)	1.02 (0.77, 1.34)

[a] Table reports prevalence for SA 1 and adjusted incidence otherwise. Results are from weighted logistic regressions with robust variance estimation.
[b] Outcome ascertainment is restricted to subjects' first screening episode only.
[c] Outcome ascertainment period is expanded to include all screenings spaced apart by >9 months (404,558 person-screenings in total; see Figure 2.2).
[d] Vaccine exposure level is assigned according to the number of vaccine doses received prior to 30, 90, or 180 days before the first screening. The rationale is to exclude doses received shortly before the start of outcome ascertainment to account for HPV infections that were likely to pre-exist those vaccine doses.
SA, sensitivity analysis; CIN, cervical intraepithelial neoplasia.
**p<0.05; **p<0.01; ***p<0.001*

Table 2.5: Subgroup analyses: Adjusted incidence of any cytological abnormality

Stratification variable	N	Adjusted incidence rate (per 100 annual person-screenings), by dose level[a]				Adjusted relative risk (95% CI)		
		0	1	2	3	1 vs. 0	2 vs. 0	3 vs. 0
Chlamydia or gonorrhea testing								
Yes	133,438	12.85	11.42	11.01	9.91	0.87(0.82, 0.93)***	0.84(0.80, 0.89)***	0.75(0.71, 0.78)***
No	101,391	10.09	9.71	8.62	8.00	0.96(0.89, 1.04)	0.84(0.78, 0.90)***	0.77(0.73, 0.82)***
Median family income (4-person)								
≤200% FPL	21,664	13.69	12.44	11.31	11.08	0.90(0.77, 1.04)	0.80(0.69, 0.93)**	0.79(0.69, 0.90)***
201-300% FPL	60,892	12.55	11.52	10.23	9.37	0.91(0.83, 1.00)*	0.79(0.72, 0.87)***	0.72(0.67, 0.78)***
301-400% FPL	65,843	11.37	10.81	10.40	8.83	0.94(0.86, 1.03)	0.90(0.84, 0.98)*	0.76(0.71, 0.81)***
>400% FPL	81,389	10.51	9.47	9.06	8.51	0.89(0.82, 0.97)**	0.85(0.79, 0.91)***	0.79(0.75, 0.83)***
Quartiles, % adults without high school completion								
1	73,848	10.58	9.80	8.96	8.32	0.92(0.84, 1.00)*	0.83(0.78, 0.89)***	0.77(0.73, 0.81)***
2	59,167	11.42	10.60	10.06	8.85	0.92(0.84, 1.01)	0.87(0.80, 0.94)***	0.75(0.70, 0.81)***
3	55,470	12.04	11.73	10.68	9.39	0.97(0.88, 1.07)	0.87(0.80, 0.96)**	0.76(0.70, 0.82)***
4	41,303	13.08	10.92	10.65	10.28	0.81(0.72, 0.92)***	0.79(0.71, 0.88)***	0.76(0.69, 0.84)***
Quartiles, % white race								
1	44,432	13.36	12.06	11.87	10.35	0.89(0.80, 0.98)*	0.87(0.79, 0.96)**	0.75(0.69, 0.82)***
2	60,133	11.44	10.54	9.56	9.15	0.91(0.83, 1.00)	0.82(0.75, 0.89)***	0.78(0.73, 0.84)***
3	66,930	11.43	10.64	9.74	8.59	0.92(0.84, 1.01)	0.84(0.77, 0.90)***	0.73(0.68, 0.78)***
4	58,293	10.64	9.81	9.21	8.53	0.91(0.82, 1.01)	0.85(0.78, 0.93)***	0.78(0.73, 0.84)***
Outpatient service use in 2007								
Yes	204,138	11.51	10.68	9.74	9.04	0.92(0.87, 0.97)***	0.83(0.79, 0.87)***	0.76(0.74, 0.79)***
No	30,691	12.20	10.31	11.57	9.12	0.83(0.71, 0.97)*	0.94(0.80, 1.11)	0.72(0.62, 0.85)***

[a] Results are from weighted pooled logistic regression models with person-screening as the unit of analysis. The weighted models include indicators for vaccine dose level, the subgroup of interest, interactions between the subgroup and dose level, and screening episode number. The finding of no effect modification is based on formal tests for the overall significance of the interaction terms between categories of dose level and of each stratification variable (all p>0.05).

**p<0.05; **p<0.01; ***p<0.001*

Table 2.6: Subgroup analyses: Adjusted rates of histology-confirmed high-grade CIN by dose level

Stratification variable	N	Adjusted incidence rate (per 100 annual person-screenings), by dose level[a]				Adjusted relative risk (95% CI)		
		0	1	2	3	1 vs. 0	2 vs. 0	3 vs. 0
Chlamydia or gonorrhea testing								
Yes	133,438	0.78	0.51	0.50	0.35	0.65(0.51, 0.84)***	0.63(0.50, 0.80)***	0.45(0.36, 0.55)***
No	101,391	0.63	0.35	0.48	0.32	0.54(0.38, 0.77)***	0.76(0.57, 1.00)	0.50(0.38, 0.65)***
Median family income (4-person)								
≤200% FPL	21,664	0.69	0.54	0.44	0.31	0.79(0.40, 1.58)	0.64(0.34, 1.22)	0.45(0.26, 0.78)**
201-300% FPL	60,892	0.79	0.41	0.53	0.35	0.52(0.35, 0.78)**	0.67(0.46, 0.99)*	0.45(0.31, 0.65)***
301-400% FPL	65,843	0.81	0.52	0.60	0.38	0.64(0.45, 0.91)*	0.74(0.54, 1.02)	0.47(0.35, 0.63)***
>400% FPL	81,389	0.60	0.35	0.38	0.29	0.59(0.40, 0.87)**	0.64(0.48, 0.84)**	0.49(0.38, 0.64)***
Quartiles, % adults without high school completion								
1	73,848	0.63	0.33	0.42	0.32	0.53(0.34, 0.80)**	0.66(0.48, 0.92)*	0.50(0.39, 0.65)***
2	59,167	0.77	0.58	0.54	0.30	0.76(0.53, 1.08)	0.70(0.51, 0.96)*	0.39(0.29, 0.53)***
3	55,470	0.78	0.48	0.53	0.40	0.61(0.42, 0.90)*	0.68(0.47, 0.99)*	0.51(0.36, 0.74)***
4	41,303	0.71	0.34	0.49	0.33	0.48(0.26, 0.87)*	0.69(0.43, 1.12)	0.46(0.28, 0.74)**
Quartiles, % white race								
1	44,432	0.64	0.52	0.42	0.25	0.82(0.53, 1.27)	0.66(0.43, 1.03)	0.40(0.25, 0.64)***
2	60,133	0.69	0.42	0.46	0.40	0.61(0.40, 0.92)*	0.66(0.47, 0.92)*	0.57(0.42, 0.77)***
3	66,930	0.80	0.40	0.48	0.36	0.49(0.34, 0.71)***	0.60(0.43, 0.84)**	0.45(0.33, 0.61)***
4	58,293	0.69	0.43	0.58	0.30	0.62(0.39, 0.96)*	0.84(0.59, 1.19)	0.44(0.31, 0.61)***
Outpatient service use in 2007								
Yes	204,138	0.73	0.47	0.49	0.34	0.64(0.52, 0.79)***	0.67(0.56, 0.80)***	0.46(0.39, 0.55)***
No	30,691	0.65	0.21	0.51	0.34	0.33(0.13, 0.82)*	0.78(0.36, 1.71)	0.52(0.24, 1.11)

[a] Results are from weighted pooled logistic regression models with person-screening as the unit of analysis. The weighted models include indicators for vaccine dose level, the subgroup of interest, interactions between the subgroup and dose level, and screening episode number. The finding of no effect modification is based on formal tests for the overall significance of the interaction terms between categories of dose level and of each stratification variable (all p>0.05).
**p<0.05; **p<0.01; ***p<0.001*

Table 2.7: Adjusted cumulative incidence of Pap screening initiation by number of HPV vaccine doses

Age in 2007	N	Cumulative percent of subjects initiating Pap screening							
		By end of 2010				By end of 2013			
		0 dose	1 dose	2 dose	3 dose	0 dose	1 dose	2 dose	3 dose
All ages	1,226,763	15.3%	19.1%	18.9%	19.2%	31.3%	37.2%	37.0%	37.3%
17 years	103,471	43.9%	53.3%	53.0%	53.5%	69.5%	79.1%	78.8%	79.3%
16 years	139,773	36.8%	45.4%	45.1%	45.6%	66.7%	76.5%	76.2%	76.7%
15 years	146,971	28.0%	35.1%	34.9%	35.3%	58.4%	68.5%	68.2%	68.8%
14 years	147,230	17.8%	22.8%	22.6%	22.9%	41.8%	50.9%	50.6%	51.2%
13 years	145,215	9.4%	12.3%	12.1%	12.3%	23.4%	29.6%	29.4%	29.8%
12 years	142,389	4.5%	5.9%	5.9%	5.9%	14.6%	18.7%	18.6%	18.8%
11 years	137,979	1.7%	2.3%	2.3%	2.3%	7.8%	10.1%	10.0%	10.2%
10 years	134,492	0.4%	0.6%	0.6%	0.6%	3.3%	4.3%	4.3%	4.3%
9 years	129,243	0.1%	0.1%	0.1%	0.1%	1.3%	1.7%	1.7%	1.7%

Notes: Cumulative incidences of screening by dose level are estimated from the birth year-stratified Cox proportional hazards model shown in Appendix B.3, based on the average values of all other covariates. Because vaccination status is time-varying in the model, Table 2.7 conservatively assumes that all vaccine doses were delivered before January 1, 2008, which maximizes the divergence between the dose groups.

Table 2.8: Sensitivity analysis using alternative study design: Time from HPV vaccine initiation until abnormality detection among age-eligible female enrollees who initiated vaccination while screening-naïve

a. With right censoring at the earlier of: disenrollment; or an 18-month gap without another screening episode after an individual has initiated screening

Endpoint / Comparison	Adjusted hazard ratio (95% CI)[a] [Sensitivity analysis with alternative cohort design]			Adjusted relative risk (95% CI)[b,c] [Main analysis]		
Any cytological abnormality						
3 vs. 1 doses	0.77	(0.74, 0.81)	***	0.84	(0.79, 0.88)	***
3 vs. 2 doses	0.86	(0.82, 0.90)	***	0.90	(0.86, 0.95)	***
High-grade CIN (CIN2+)						
3 vs. 1 doses	0.70	(0.55, 0.88)	**	0.77	(0.60, 0.99)	*
3 vs. 2 doses	0.69	(0.56, 0.84)	***	0.69	(0.54, 0.87)	**

b. With right censoring at the earlier of: disenrollment; or end of the first cervical screening episode

Endpoint / Comparison	Adjusted hazard ratio (95% CI)[a] [Sensitivity analysis with alternative cohort design]			Adjusted relative risk (95% CI)[c,d] [Sensitivity analysis 1 from Table 2.4]		
Any cytological abnormality						
3 vs. 1 doses	0.83	(0.78, 0.88)	***	0.82	(0.77, 0.88)	***
3 vs. 2 doses	0.93	(0.88, 0.98)	**	0.92	(0.86, 0.98)	**
High-grade CIN (CIN2+)						
3 vs. 1 doses	0.70	(0.51, 0.94)	*	0.69	(0.50, 0.95)	*
3 vs. 2 doses	0.70	(0.53, 0.92)	*	0.65	(0.48, 0.88)	**

[a] The multivariate analyses of months from vaccine initiation to abnormality detection include the 411,312 age-eligible enrollees who initiated vaccination prior to screening or disenrollment (whichever occurred first). Vaccination status is a time-varying variable reflecting the number of doses received up to and including each month of follow-up. (Due to endogeneity concerns, vaccine doses received at or after the first Pap screening visit do not count towards an individuals' vaccination status.)
[b] From the main analysis in Table 2.3. The weighted pooled logistic regression analysis of abnormalities per annual person-screening is restricted to the screening cohort (N=234,829), which includes 87,488 subjects with ≥1 dose prior to screening.
[c] Hazard probability odds ratios are taken as approximations of the relative risks comparing different dose levels.
[d] From sensitivity analysis 1 in Table 2.4, which focuses on the first screening episode only. As with the main analysis, this analysis is restricted to the screening cohort (N=234,829), which includes 87,488 subjects with ≥1 dose prior to screening.
**p<0.05; **p<0.01; ***p<0.001*

Comparisons with prior epidemiologic data

For CIN1, CIN2+, and CIN3+, risk reductions with three (versus zero) doses in the present study are close to the estimated proportions of CIN1, CIN2+, and CIN3+ cases attributed to HPV-6/11/16/18 in the U.S. (Table 2.9).

Table 2.9: Estimated reductions in CIN with three doses in present study, alongside reported proportions of CIN attributable to HPV-6/11/16/18 in U.S.

Endpoint	% Reduction in present study (3 vs. 0 doses)[a]		Estimated proportion of cases attributed to HPV-6/11/16/18 in United States[b]	
	Main analysis (95% CI)	Range in sensitivity analyses	Insinga (2008)[c]	Hariri (2012)[d]
CIN1	20% (14, 25)	18-23%	20%	-
CIN2+	53% (45, 60)	50-55%	59%	54%
CIN3+	67% (54, 77)	58-70%	-	66%

[a] Percent reductions, calculated as (1 - relative risk)*100%, are based on relative risks from the main analysis (Table 2.3) and sensitivity analyses (shown in Table 2.4 for the primary study endpoints). For CIN1, I refer to the percent reductions corresponding to low-grade CIN in my analyses.
[b] Attributable proportions are among pre-invasive cervical lesions only (i.e., highest-severity cases are CIN3 or adenocarcinoma in situ, not invasive cervical cancer).
[c] The meta-analysis by Insinga (2008) estimated the attribution of individual HPV types in CIN1 and CIN2/3, with adjustment for the co-occurrence of HPV types.
[d] Hariri (2012) estimated HPV type attributions in a sample of U.S. women aged 18-39 years within the HPV-IMPACT monitoring system who had a diagnosis of CIN2+ during 2008-2009. The authors noted that these estimates may have been somewhat impacted by HPV vaccination in a small percentage of the sample.

Appendix B.4 summarizes vaccine efficacy against all-cause CIN (i.e., irrespective of HPV type) from quadrivalent and bivalent trials among 15-26 year-old women. For all-cause CIN2+, my effect estimate for three versus zero doses (53%; 95% CI: 45-60) is larger than efficacy within the ITT trial populations, but comparable to efficacy within the HPV-naïve populations from the quadrivalent (43%; 95% CI: 24-57) and bivalent (65%; 95% CI: 53-74) trials (Muñoz 2010; Lehtinen 2012).

The comparison of effect sizes for all-cause CIN3+ follows a similar pattern as CIN2+. Conversely, for all-cause CIN1, I observe reductions of 20% (95% CI: 14-25) with three versus zero doses, which is closer to the reported efficacy of the quadrivalent vaccine in the ITT population (20%; 95% CI: 12-28) than in the HPV-naïve population (30%; 95% CI: 17-41) (Muñoz 2010).

DISCUSSION

Key findings and policy implications

To my knowledge, this study represents the largest and the first U.S.-based analysis of cervical abnormalities by number of HPV vaccine doses. Completion of the three-dose series prior to screening is associated with significantly lower rates of all abnormality endpoints. Receipt of one or two doses is associated with significant but smaller decreases across most endpoints. For all dose levels, risk reductions with vaccination are greatest with respect to CIN3+, as expected given that the proportion of abnormalities caused by HPV-16/18 increases with greater lesion severity (Insinga 2008; Hariri 2012).

This study extends prior research describing the public health impact of HPV vaccination. Ecological data have shown large declines in HPV infections and genital warts following vaccine introduction in the U.S (Bauer 2012; Kahn 2012; Markowitz 2013). An individual-level analysis of national survey data observed 82% lower HPV-16/18 prevalence among vaccinated (≥1 dose) versus unvaccinated sexually active adolescent females (Markowitz 2013). Outside of the U.S., a population-based cohort study in Sweden reported monotonic decreases in genital wart incidence with each additional dose of the quadrivalent series (Herweijer 2014). Conversely, in an analysis of cervical abnormalities among Australian 11-27

year-olds in 2007, modest risk reductions versus no vaccination were seen with one dose, but the estimated reductions were only statistically significant starting from the two-dose level (Crowe 2014). Initiation/completion rates were high (84%/70%) within the region's school-based vaccination program, so the lack of significance with one dose may have been due to relatively small numbers (Crowe 2014).

Internationally, there has been growing interest in one- or two-dose HPV vaccination schedules (Markowitz 2014; President's Cancer Panel 2013; Natunen 2011; WHO 2014). In settings where the costs or logistics of three-dose vaccination are prohibitive, effectiveness evidence on fewer than three doses could be a critical factor for country-level vaccine adoption decisions (Natunen 2011; WHO 2014). In the U.S., the recommended three-dose regimen is unlikely to change on the basis of population-based surveillance studies, especially if future studies continue to show optimal risk reductions with three doses. However, observational data could guide policy and clinical efforts to increase HPV vaccine coverage, considering that the goals of increasing vaccine initiation versus improving multi-dose adherence probably require different strategies. Real-world evidence of significant protection with one or two doses may also encourage better vaccine acceptability at the individual level in both high- and low-resource settings (Kahn 2008; Natunen 2011; Francis 2010).

To date, trial-based evaluations of reduced-dose regimens have been restricted to intermediate endpoints. The available data suggest that partial vaccination may be efficacious but also provide some evidence of incremental benefit from a third dose. In a Canadian trial of 9-13 year-old girls randomized to two or three quadrivalent vaccine doses, antibody responses to HPV-6/11/16/18 were non-inferior with two doses at month 7, but non-inferiority was lost at subsequent visits for HPV-6/18 (Dobson 2013). In a recent post-hoc analysis of the Costa Rica

and PATRICIA trials, one, two, and three doses of bivalent vaccine showed similar protection against persistent 16/18-type infections over four years of observation (Kreimer 2015). Cross-protection against persistent 31/33/45-type infections was limited to three doses, although a sub-analysis suggested that cross-protection may occur with two doses spaced apart by ≥6 months (Kreimer 2015). In the Costa Rica trial, antibodies against 16/18 remained stable for four years even with one dose, but were approximately five times higher with three doses (Safaeian 2013).

My effect estimates for full vaccination are supported by data from epidemiologic studies and vaccine trials among 15-26 year-old women. For example, the 53% (95% CI: 45-60) reduction in CIN2+ with three doses aligns with the U.S. proportion of CIN2+ cases attributed to HPV-16/18 (54-59%) (Insinga 2008; Hariri 2012). Assuming no cross-protection, this 16/18-attributable proportion provides an estimate for the percent of CIN2+ that is vaccine-preventable, given that the vaccines prevented 95-98% of *16/18-related* CIN2+ when administered before infection (Lehtinen 2012; Kjaer 2009). Vaccination only prevented 52-61% of 16/18-related CIN2+ within the mix of HPV-exposed and –unexposed trial participants; however, mean age at vaccine initiation was approximately four years older than in the present study.

Limitations

Strengths of this study include the use of a large administrative database with enrollees in all U.S. states, which provides sufficient sample and follow-up for a dose-stratified analysis of cervical endpoints. However, this study is subject to some important limitations. Because it is observational, bias from unmeasured confounders is possible. Within the study sample, ZIP code-level socioeconomic advantage tends to increase with increasing dose level. Other unobserved factors could correlate with both vaccination status and HPV exposure risk.

Nevertheless, I use IPTW to control for a number of covariates likely to be associated with vaccine exposure and the outcomes. This approach yields balanced distributions of observed covariates across the dose groups, and does not appreciably change my estimates. The magnitude and pattern of effect sizes are also well aligned with expectations for vaccines targeting HPV-16/18.

Indeed, one key threat to validity is that sexual history and HPV vaccination status could be correlated. My analyses include reasonable proxies for higher HPV exposure risk (i.e., history of testing/diagnoses for other sexually transmitted infections), which do not monotonically or substantially differ across dose groups. Previous survey studies also have found no significant associations between HPV vaccination status and age of sexual initiation (Markowitz 2013; Liddon 2012; Marchand 2013) or number of partners (Liddon 2012; Marchand 2013), though one study noted more sex partners among vaccinated (versus unvaccinated) adolescents (Markowitz 2013).

Some misclassification of exposure could occur due to vaccinations not captured by administrative claims. Yet such underreporting should be minimal because HPV vaccines are expensive and covered by private insurance. Further, few study subjects would have been eligible for federally funded free vaccinations.

Outcomes data are unavailable for age-eligible enrollees who did not initiate screening during the study timeframe; however, I expect those who underwent screening to be a more policy-relevant population of adolescents who were likely past sexual initiation by the time of outcome ascertainment. Sensitivity analyses suggest that potential selection bias from conditioning on screening participation is unlikely to account for the observed dose-response patterns, particularly among the ≥1 dose levels. Screening-related selection bias also would not

explain the analogous finding in Sweden of a dose-response relationship between HPV vaccination and genital warts, a symptomatic rather than screening-detected endpoint (Herweijer 2014).

Despite its size, MarketScan is not representative of the entire private insurance market or general U.S. population. While I do not observe effect modification across socioeconomic subgroups, vaccine effectiveness may be sensitive to population-specific characteristics such as timing of sexual initiation and prevalence of vaccine-targeted HPV types.

Conclusions

Adolescent girls who received three HPV vaccine doses show significantly reduced risks of cervical abnormalities, including high-grade lesions that are considered precursors to cervical cancer. Partial vaccination with one or two doses is associated with significant but lesser risk reductions. These results, combined with preliminary trial evidence, suggest that HPV vaccination may be reasonably protective even when incomplete. My findings highlight the importance of increasing vaccine uptake, even when the patient or provider anticipates difficulty ensuring completion of subsequent doses. Further investigation is needed to assess both the efficacy and durability of protection from partial vaccination. Evidence of significant protection with only one or two doses could be an important driver of HPV vaccine acceptance, accessibility, and affordability in many countries.

Paper 3:

Cost-effectiveness of performance-based contracts for obstetric care providers in rural India

INTRODUCTION

Maternal mortality and provider quality in India

Globally, there were an estimated 303,000 deaths due to pregnancy and childbirth in 2015, with the vast majority occurring in developing countries (WHO 2015a). Maternal deaths usually occur during the labor, delivery, and immediate postpartum stages of maternity (Lassi 2014), and are largely preventable using available and effective medical interventions. Postpartum hemorrhage, the single largest cause of maternal mortality, can often be averted through inexpensive prophylactic measures (e.g., administration of oxytocic drugs during the third stage of labor, uterine massage); meanwhile, interventions such as blood transfusions, further oxytocic treatment, or manual removal of the placenta can effectively manage bleeding in time to prevent death (WHO 2012; Hofmeyr 2013). Sepsis and sepsis-related mortality can be prevented through clean delivery practices, prophylactic antibiotics during caesarean section, and therapeutic antibiotics (WHO 2008). Monitoring, drug therapy, and induction of labor when appropriate can prevent the progression of pre-eclampsia to eclampsia, as well as lower the risk of death from these hypertensive disorders (WHO 2011).

In India, poor maternal health outcomes continue to be a significant problem despite large increases in hospital-based deliveries over the past decade (Das & Hammer 2014; IIPS 2014). Low-quality medical care is one likely cause of pregnancy-related complications and death among women who give birth in health facilities. Evidence suggests that poor quality of care in India stems from a combination of insufficient knowledge of clinically appropriate procedures and deficits in effort among providers (Das 2008; Das 2012a; Tielsch 2015). In a vignette study of primary care providers (in which actors posed as patients to test providers' knowledge), Das et al. (2012a) found low levels of adherence to basic essential procedures for common illnesses,

with only small differences in adherence between providers with and without medical qualifications. Another study that combined vignettes with direct observation of patient-provider interactions noted a gap between what trained providers know versus what they do: In Delhi's private sector, providers without an MBBS knew only 20% of essential tasks but performed nearly all of this 20%, whereas providers with an MBBS knew 40% of essential tasks when tested but performed only 25% (Das 2008). Reward payments based on performance measures may help address such "know-do" gaps by aligning provider incentives with patient health goals.

Cost-effectiveness of paying for quality: State of the evidence

Under pay-for-performance (P4P), health care providers receive financial incentives based on pre-specified performance indicators, which may relate to process-of-care quality or (less commonly) patients' health outcomes. P4P schemes are now widespread in the U.S. and other high-income countries (Eijkenaar 2012), and have become increasingly popular in low- and middle-income countries as a policy instrument to encourage the delivery of high-value medical interventions (Witter 2012; Meessen 2011; Barter 2014; Miller 2013). Research on P4P has yielded mixed results but suggests that P4P programs can be effective in achieving quality improvements depending on the health context and specific design features (e.g., whether rewards are allocated to individual providers, teams, or organizations; type of performance measures used; secondary program features such as training or feedback, etc.) (Eijkenaar 2013; Miller 2013).

However, it remains unclear whether P4P represents a good use of resources, even when it is found to have the desired effects on quality. Most evaluations of P4P initiatives have focused exclusively on their effectiveness without considering costs or cost-effectiveness, an

evidence gap that has been highlighted by several review articles and commentaries (Maynard 2012; Meacock 2014; Eijkenaar 2013; de Bruin 2011; Emmert 2012; Peterson 2006). Past economic evaluations of P4P have generally been inconclusive because they either failed to incorporate a sufficiently broad range of cost categories, or did not attempt to convert measures of P4P program effectiveness into standard metrics (e.g., quality-adjusted life years) that would enable comparisons with recommended willingness-to-pay thresholds (Emmert 2012). Cost-effectiveness evidence is especially sparse in resource-constrained settings, where operating efficient P4P programs may be particularly challenging (e.g., due to a lack of pre-existing infrastructure for collecting performance data) (Borghi 2015).

Specific aims

My paper examines the cost-effectiveness of two P4P interventions that aim to improve maternal health in rural Karnataka, India. The cost-effectiveness model uses primary data from a recently-completed randomized experiment, which tested two forms of P4P versus a control arm among private obstetric care providers: (i) outcome-based P4P, where rewards are based on the measured occurrence of obstetric complication outcomes; and (ii) input-based P4P, where rewards are based on measures of adherence to guideline-recommended practices. In this study, the input-based approach appeared to generate a 9 percentage point reduction in the probability of postpartum hemorrhage compared to the control arm, although similar improvements were not seen with respect to other measured complications or under outcome-based P4P. I extrapolate from these trial findings to assess whether the interventions are economically attractive given region-specific willingness-to-pay thresholds.

METHODS

Overview of randomized experiment

Participants

The Experimental Evaluation of Performance Incentive Contracts was a cluster-randomized experiment designed to evaluate the effectiveness of P4P contracts aiming to improve the quality of maternal health care in rural Karnataka (Mohanan 2015). The study was jointly funded by the State Government of Karnataka and external donors, including the World Bank, the International Initiative for Impact Evaluation, and the UK Department for International Development.

Participants were eligible if they were private medical providers who performed obstetric services in rural areas of the state, defined as *hoblis* in which there were no large public health providers (e.g., district hospital, taluk hospital, sub-divisional hospital, community health center, or 24/7 primary health center). Providers were ineligible if they provided obstetric care infrequently (<24 deliveries per year), or moved from the region before randomization. Providers were recruited into the study through a three-stage process that included field work seeking to identify all formal obstetric care providers in targeted regions of Karnataka, interviews with providers to assess eligibility, and snowball sampling to identify additional eligible providers missed by the initial field work. Details of the recruitment and screening process are presented in Figure 3.1.

Financial incentives

Providers were randomly assigned to one of three intervention arms: input-based P4P, outcome-based P4P, or control contracts.

The *outcome-based P4P* arm received financial rewards contingent on the measured risk of four adverse maternal and neonatal health outcomes among their patients: postpartum hemorrhage, pre-eclampsia/eclampsia, sepsis, and neonatal mortality. For each maternal health outcome, providers were rewarded per percentage point reduction in the estimated risk of the complication that they achieved in their patient population relative to a pre-specified performance benchmark. The reward payment to provider *p* for outcome *i* was calculated as:

$$R(y_{ip}) = \begin{cases} a_i(\overline{y}_i - x_{ip}) * 100, & \text{if } y_{ip} \leq \overline{y}_i \\ 0 & \text{, if } y_{ip} > \overline{y}_i \end{cases}$$

where y_{ip} is the measured probability of outcome *i* in the provider's patient population; $\overline{y}_i$ is the estimated pre-intervention probability of outcome *i* for the average provider in rural Karnataka ($\overline{y}_{pph}$=0.35, $\overline{y}_{pe}$=0.20, and $\overline{y}_s$=0.08); and a_i is a constant corresponding to the incremental reward (in Rs.) per percentage point reduction, which was predetermined based on available budget and the anticipated range of provider performance for outcome *i* (a_{pph}=850, a_{pe}=1750, and a_s=8650). For neonatal death, providers under outcome P4P were given a fixed reward of Rs. 15,000 if they achieved zero neonatal deaths in their population.

The *input-based P4P* arm instead received rewards based on measures of adherence to clinically appropriate processes of care among their patients. Process measures were formulated using World Health Organization (WHO) guideline recommendations for basic obstetric care (WHO 2009) and included five domain scores: Pregnancy Care, Childbirth Care, Postnatal Maternal Care, Newborn Care, and Postnatal Newborn Care. Domain scores were each calculated on a 0 to 1 scale, based on a simple average of binary indicators for specific within-domain health service inputs (2 to 26 indicators, depending on the domain). Rewards were

calculated per percentage point improvement in each domain score versus a pre-specified performance benchmark. For each domain *k*, the reward payment to provider *p* was:

$$R(x_{kp}) = \begin{cases} b_k(x_{kp} - \overline{x}_k) * 100, & \text{if } x_{kp} \geq \overline{x}_k \\ 0 & \text{, if } x_{kp} < \overline{x}_k \end{cases}$$

where x_{kp} is the measured score for domain *k* in the provider's patient population; $\overline{x}_k$ is the estimated pre-intervention average score for domain *k* among providers in rural Karnataka ($\overline{x}_{pc}$=0.85, $\overline{x}_{cc}$=0.65, $\overline{x}_{pmc}$=0.50, $\overline{x}_{nc}$=0.80, and $\overline{x}_{pnc}$=0.70); and b_k is a constant corresponding to the reward (in Rs.) per percentage point improvement (b_{pc}=3700, b_{cc}=750, b_{pmc}=450, b_{nc}=1850, and b_{pnc}=950).

Providers in the input- and outcome-based P4P arms were informed that inputs or outcomes would be measured through household surveys of patients who present to them for delivery over a 1-year period. They were told that they could potentially earn up to a maximum of roughly Rs. 150,000 (~$2,500 at the time, which is over 15% of a mid-level doctor's salary). To minimize the likelihood that providers selectively turn away high-risk patients, their contracts stated that the reward payment would be voided if there is evidence in the local population of refusal to provide care (with the exception of medically appropriate referrals to higher-tier facilities). Contracts included a similar provision that rewards would be voided if it is found that providers selectively reported deliveries that took place in their facility.

Providers in the *control contracts* arm were offered no financial incentives for performance, but were told about the input and outcome measures and provided with the same educational materials as the other arms (the 2009 WHO guidelines and a detailed set of guidelines from the Government of India). Additionally, all study arms received three

installments of Rs. 2,500 (~$41) for their participation in the program and another payout of Rs. 1,000 (~$16) for their efforts in transmitting patient lists.

Study procedures

Study procedures included a total of four in-person provider visits. During the baseline visit (Oct 2012 – Jan 2013), the field team collected preliminary data about providers, including their volume of deliveries and facility capabilities. At the first intervention visit (February-April 2013), providers received educational materials and were randomized to an intervention arm after an initial series of questions to confirm eligibility (Figure 3.1). The second intervention visit (May-August 2013) included an open-ended interview with providers to discuss any quality improvement efforts they had undertaken and other topics. At this visit, the team also arranged for biweekly collection of patient lists from the provider. At the third intervention visit (September-November 2014), which occurred after the household data collection was complete, the team collected follow-up data on facility capabilities and supplies and distributed rewards.

Performance data was collected via household surveys of patients chosen at random from providers' full patient lists over the course of a year from the second intervention visit. Interviews were conducted at approximately two weeks from the delivery to prevent loss of recall (Das 2012b). The main household instrument was an approximately 1-hour closed-ended questionnaire and included questions pertaining to complication outcomes, input quality, and a range of other topics (e.g., demographic and socioeconomic information, detailed birth history, general health history, total hospital expenditure for delivery at the provider's facility, etc.).[8]

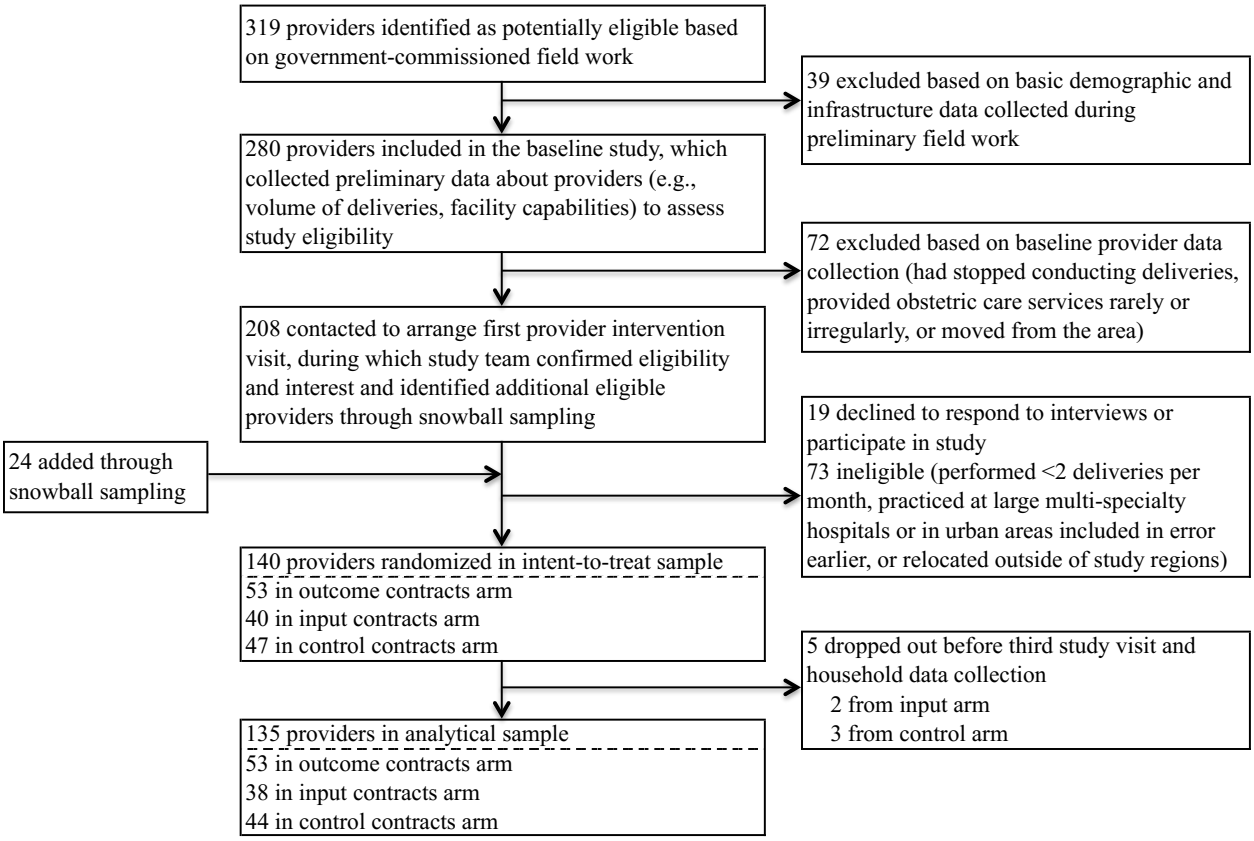
[8] The specific questions used to measure obstetric complications were formulated based on published validation studies by Filippi et al. (2000), Stewart & Festin (1995), and Souza et al. (2010). Because no studies had been conducted to validate process-of-care measures of maternal health care quality based on patient recall, study investigators undertook such a validation study in Gujarat and Karnataka prior to the experiment (forthcoming). The

The intent-to-treat population included the 140 providers randomized during the first intervention visit (Figure 3.1). Attrition was low and not statistically different between the groups. Overall, five providers opted out of the study prior to household data collection (two from the input group, three from the control group). The final analytical sample thus included 135 providers, with a corresponding household survey sample of 2,608 patients. To reduce the potential for bias due to provider dropout after randomization, the main estimates of program effect in the impact evaluation study and the present cost-effectiveness analysis are risk-adjusted using a pre-specified set of patient- and provider-level characteristics.

---

study used senior nursing students to observe and document whether various indicators of input quality were provided, and these data were subsequently compared to responses from interviews with the same mothers approximately two weeks later. Measures with high sensitivity and specificity were selected for use during the randomized experiment.

Figure 3.1: Flowchart of participating obstetric care providers



Model overview

Using a decision tree approach, the cost-effectiveness model simulates a heterogeneous cross-sectional population of pregnant women presenting for delivery at various participating provider facilities during the 1-year contract period. The modeling timeframe spans each individual's remaining lifetime, counting from her age at the contract year delivery. The hypothetical population is constructed to mirror characteristics in the target population of patients and providers in rural Karnataka. Specifically, for each woman included in the original household survey sample, the model population includes a clone with identical baseline characteristics (e.g., demographics, obstetric history, and pre-contract provider-level factors), but

with randomization group reassigned according to the policy scenario being modeled: (i) status quo; (ii) input-based P4P; or (iii) outcome-based P4P.

Within the 1-year contract period, women's risks of delivery-related complications vary depending on both the policy scenario and patient/provider characteristics; I derive these risk equations through primary analyses of the trial data. Women who experience complications are at risk of maternal mortality, conditional on the type of complication(s). P4P interventions are assumed to have no impact on the case fatality rate for a given complication type. Beyond the contract period, those who survive the delivery in Year 1 are assumed to live out their remaining life expectancy, with future life-years discounted at a rate of 3% per year. Costs are calculated from both a societal and program perspective and include average costs per delivery associated with program implementation (e.g., costs of performance and participation rewards, meetings with providers, data collection, etc.), the medical management of complications, and uncompensated provider effort. Health outcomes are similarly expressed per patient delivery and include the expected number of maternal complications, number of deliveries with ≥1 complication(s), number of maternal deaths (overall and by cause), and discounted life-years. Incremental cost-effectiveness ratios (ICERs) are calculated as cost per life year saved, as well as cost per maternal complication averted. As recommended by the WHO, I consider interventions with ICERs (cost per life year saved) less than gross domestic product (GDP) per capita to be very cost-effective, and those with ICERs less than three times GDP per capita to be reasonably cost-effective (WHO 2015b).

Model calculations are deterministic and are performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). Because expected health outcomes are subject to patient heterogeneity and are calculated through a non-linear function of patient- and provider-level

characteristics, the model first performs a separate deterministic calculation for each individual in the hypothetical population and then takes the average of expected outcomes and costs across individuals (Koerkamp 2011).

Parameter estimation

Risk of complications and intervention effectiveness

The model assumes that P4P interventions affect life expectancy within the target population solely by changing the occurrence of major obstetric complications that cause maternal mortality. Based on the set of maternal complications measured in the household survey, I consider three adverse clinical outcomes of delivery[9] (postpartum hemorrhage, sepsis, and pre-eclampsia/eclampsia) and assume that P4P interventions have no impact on other unmeasured complication types.

Using household survey data from the experiment, I estimate the following logistic regression equation to model the absolute risk of each obstetric complication among women who present to participating providers for delivery, conditional on treatment arm and exogenous risk factors:

$$\text{logit } P[y_{ip}=1|T_p, X_p, Z_i] = B_0 + B_1T_p + B_2X_p + B_3Z_i$$

Above, y_{ip} is an indicator for the complication of interest for patient i seeking delivery assistance from provider p, and T_p is a vector of treatment group indicators. X_p is a vector of provider characteristics measured before the contract period, including: gender; professional qualifications

[9] Given the non-differential risk of neonatal mortality across the three randomization arms (Mohanan 2015) and the general convention of not incorporating neonatal health outcomes in decision models of maternal health (Goldie 2010), neonatal death is not considered as a model outcome in the cost-effectiveness analysis. However, program expenditures under outcome P4P include the Rs. 15,000 reward given to providers who achieve zero neonatal deaths (a reward which nearly all providers received due to the rarity of this outcome).

(MBBS, BAMS, or other); number of years in practice; and district where the provider's facility is located. Z_i is a vector of the following pre-delivery or time-invariant patient characteristics: mother's age (in years) and age^2 at delivery; history of hypertension, diabetes, and stomach surgery; whether it is the mother's first pregnancy; whether the mother has previously had a stillbirth or abortion; number of previous children birthed; mother's education level (illiterate; attended primary school only; attended secondary school or higher); household's caste (scheduled caste; scheduled tribe; other backward class; general class or other unspecified caste); whether the household owns land; and whether the household owns a Below Poverty Line card (i.e., an identifier issued by the state government to households eligible for various assistance programs).[10]

The cost-effectiveness model uses the fitted risk equations to obtain predicted probabilities of obstetric complications for each woman in the hypothetical cohort under each model scenario. I conservatively assume that the adjusted complication probabilities corresponding to control arm assignment are representative of the status quo scenario, which assumes no incremental benefit from the provider intervention visits, provision of educational materials, and participation rewards alone.

Table 3.1 below presents effect estimates and adjusted outcome probabilities from each risk equation. (Detailed regression output is provided in Appendix C.1.) The odds of postpartum hemorrhage are 39% lower under input P4P compared to control contracts, which translates to a

[10] This model specification is chosen for consistency with the pre-specified full model used in the main impact evaluation, but with the following modifications to better suite the needs of the cost-effectiveness analysis: First, I fit logistic rather than linear regression equations to ensure that predicted probabilities are non-negative for all individuals in the hypothetical model population. Second, to improve the stability of the fitted equations, I remove indicators for comorbidities with low prevalence in the sample (hypo- or hyperthyroidism, asthma) and covariates that are closely correlated with other included variables (number of years that the provider's facility has been in practice, number of previous pregnancies, whether the household has no literate adults, and house type). Third, I add a quadratic term for the mother's age to reflect the possibility that both young and advanced maternal age could be risk factors for a given maternal complication.

9.6 percentage point reduction in absolute risk ($p<0.05$). Conversely, postpartum hemorrhage risk is similar between the outcome P4P and control groups. Risks of pre-eclampsia/eclampsia and sepsis do not significantly differ across the treatment groups, but are numerically higher under input and outcome P4P compared to control contracts.

Table 3.1: Impact of provider incentives on maternal complication probabilities

	Postpartum hemorrhage	Pre-eclampsia/ eclampsia	Sepsis
Odds ratio (95% CI)			
Input P4P	0.610 (0.30, 0.99)	1.152 (0.52, 2.55)	1.560 (0.61, 3.30)
Outcome P4P	0.949 (0.59, 1.59)	1.302 (0.69, 2.89)	1.654 (0.87, 3.26)
Control	(ref)	(ref)	(ref)
Provider-level controls	Y	Y	Y
Patient-level controls	Y	Y	Y
Adjusted risk of complication among patients presenting for delivery (%)			
Input P4P	28.2% (19.9, 36.3)	20.3% (12.6, 29.1)	8.3% (5.0, 12.5)
Outcome P4P	36.7% (30.3, 44.1)	22.1% (16.8, 29.0)	8.7% (6.3, 12.8)
Control	37.8% (30.9, 45.8)	18.3% (11.9, 25.0)	5.5% (3.7, 8.8)
N	2,608	2,608	2,608

Notes: Logistic regression estimates are based on surveyed households from the randomized experiment. Using the fitted risk equations, adjusted complication risks are obtained by predicting each patient's risk under each treatment group scenario (input P4P, outcome P4P, or control), and then averaging across all patients for each scenario. Confidence intervals are from cluster bootstrapping to account for intra-cluster correlation between respondents from the same provider, and are constructed using the percentile method. (Note that because the adjusted risks are estimated using the average of individual-level predicted outcomes, rather than using predicted outcomes for a hypothetical patient with average covariate values, the odds ratios presented above cannot be precisely retrieved using the adjusted complication risks above.)

These results are consistent with the main impact evaluation paper, in which investigators also present evidence of behavioral mechanisms that likely contributed to the observed reduction in postpartum hemorrhage risk under input P4P (Mohanan et al.; forthcoming). Namely, input P4P providers appeared to employ strategies such as uterine massage and active management of the third stage of labor more frequently than controls: Based on household survey responses, input P4P providers were 10 percentage points more likely to massage the abdomen after

delivery, and 6 percentage points more likely to administer oral or parenteral drugs to reduce bleeding. Consistent with these patient-reported findings, input P4P providers were 14 percentage points more likely to have parenteral oxytocic drugs routinely available, based on a difference-in-difference analysis of baseline and final visit data from hospital personnel interviews. Further consideration of the effectiveness findings can be found in the Discussion.

Natural disease history

Additional epidemiologic parameters are derived using published sources, including the case fatality rates associated with postpartum hemorrhage, pre-eclampsia/eclampsia, and sepsis, probability of maternal death due to other causes, and the remaining life expectancies of mothers who survive the contract year delivery.

The case fatality rate for each maternal complication is defined as the conditional probability of a complication-specific death among women who experience that complication. Case fatality rates are derived through a disaggregation of overall maternal mortality risk by primary cause, as shown in Table 3.2. Column [i] lists the reported proportion of maternal deaths attributable to each complication from the Sample Registration System (SRS) Special Survey of Deaths, a survey that conducted a detailed enquiry into 1,383 maternal deaths in India during 2001-2003 (Registrar General, India 2006). Because women in the target population are assumed to have survived up to the intrapartum (labor and delivery) stage of their maternity, I renormalize these attributable proportions after removing the estimated percentage of all maternal deaths that occur antepartum (before onset of delivery) (24.6%; Kassebaum 2014). The renormalized attributable proportions (column [ii]) are then multiplied by the overall risk of maternal death among women reaching the intrapartum stage, approximated as the maternal mortality ratio for

India in 2015 (174 maternal deaths per 100,000 live births; WHO 2015c) reduced by the percentage of maternal deaths occurring antepartum.[11] The resulting probabilities (column [iii]) approximate the risk of death from each complication in the target population under the status quo; each of these probabilities is divided by the estimated fraction of the target population that experiences the complication under the status quo (column [iv]), yielding the case fatality rate (column [v]).

Although the occurrence of other obstetric complications is not explicitly modeled, I assume that all women in the target population are at risk of maternal mortality from other complications (e.g., obstructed labor) and indirect causes (e.g., pre-existing anemia exacerbated by pregnancy) (Registrar General, India 2006). To estimate the risk of maternal mortality from other causes, I subtract the unconditional probabilities of death due to postpartum hemorrhage, pre-eclampsia/eclampsia, and sepsis (i.e., column [iii] of Table 3.2) from the overall risk of maternal mortality under the status quo (i.e., MMR*[1-0.246]).

For women who survive the contract year maternity, remaining life expectancy depends on age at the time of delivery. Using standard methods, life expectancy by age is derived from age-specific all-cause mortality rates for females in India (WHO 2013; see Appendix C.2 for calculation details and life expectancy table).

[11] To better approximate the true risk of death among women reaching the intrapartum stage, I adjust the numerator of the MMR to only include maternal deaths occurring intrapartum or postpartum. I make no adjustment to the denominator of the MMR, with the assumption that the number of live births in a given span of time is sufficiently close to the number of maternities that reach the delivery stage. Although there is slight under-counting due to deliveries that result in stillbirths, this bias is offset by slight over-counting due to multiple births (e.g., twins) from the same delivery (Riffe 2010).

Table 3.2: Derivation of the case fatality rates associated with maternal complications

Maternal complication	Attributable proportion of all maternal deaths [i]	Attributable proportion of intrapartum and postpartum maternal deaths [ii]=[i]/(1-p_{ante})	Probability of death from complication in target pop under Status Quo [iii] = [ii]*MMR*(1-p_{ante})	Probability of complication under Status Quo [iv]	Case fatality rate [v]=[iii]/[iv]
Postpartum hemorrhage	0.38	0.50	0.000661	0.378	0.001750
Pre-eclampsia/ eclampsia	0.05	0.07	0.000087	0.183	0.000476
Sepsis	0.11	0.15	0.000191	0.055	0.003451

Notes by table column:
[i] Maternal death is defined as the death of a woman while pregnant or within 42 days of termination of pregnancy irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes (WHO 2015a). The remaining 46% of maternal deaths in India are attributed to: complications of miscarriage or abortion, obstructed labor, other direct causes (e.g., embolism, ectopic pregnancy), and indirect causes (i.e., new or pre-existing health conditions aggravated by pregnancy, such as anemia in women without hemorrhage) (Registrar General, India 2006).
[ii] The proportion of maternal deaths occurring antepartum (p_{ante}) is estimated at 24.6% in the base-case analyses, based on a large meta-analysis by Kassebaum et al. (2014). The formula shown in column [ii] reflects my assumption that, in the vast majority of cases, deaths due to the three measured complications would be categorized as occurring in the intrapartum or postpartum stages. Although onset of the complication itself may occur before labor, the recommended course of action in such cases is often expedited delivery (WHO 2003).
[iii] The base-case value of the maternal mortality ratio (MMR) is 0.00174 (WHO 2015c). Column [iii] takes MMR*(1-p_{ante}) as an approximation for the probability of maternal death among pregnant women surviving to the intrapartum stage.
[iv] Complication probabilities are from primary analyses of trial data (see Table 3.1).
[v] The case fatality rate gives the probability of death due to the specified maternal complication, conditional on experiencing that complication.

Costs

Perspective and scope of costs

Costs are evaluated both from a program perspective and from a societal perspective. The program perspective includes accounting costs incurred by donors and the Government of Karnataka that are a result of P4P; as described below, program costs are further categorized as start-up costs or recurrent costs. In addition to these costs, the societal perspective also takes into account the economic costs of maternal complication management (which are typically incurred by the patient) and uncompensated provider effort.

In line with recommended practice for cost-effectiveness analyses alongside clinical trials (Drummond 2005), my cost calculations exclude costs of the randomized experiment that were purely protocol-driven (e.g., costs related to program impact evaluation and other research activities); the aim of this model is to evaluate the cost-effectiveness of each intervention under the real-world conditions that would be in effect if the intervention is implemented.

Start-up costs

I use an activity-based approach to estimate program costs (Drummond 2005), first identifying the core set of activities that comprise each P4P program and then quantifying the accounting costs associated with each activity based on actual spending during the P4P experiment (or by consulting with program staff when original financial reports are unavailable).

Start-up costs capture one-time capital expenses that were necessary to make the P4P programs operationally viable. As described in Appendix C.3, start-up activities include the development of provider and household survey instruments, identification of rural geographic areas to target within Karnataka, and preliminary field work to produce the initial list of potentially eligible providers and screen out ineligible providers. Note that, apart from the cost of investigators' time to develop survey instruments, the costs of designing the randomized experiment are considered research-related and are therefore not included. I estimate the cost of each activity based on project reports, original financial statements from contractors, or consultation with key investigators and program staff who oversaw the field work or developed the survey instruments.

In order to apportion program start-up costs at the individual patient level, these costs are divided out by the estimated total number of women who seek delivery assistance from

participating providers during the 1-year contract period. The number of participating providers is assumed to be 135 based on final sample size in the P4P experiment. The average number of deliveries per provider during the year (base-case value: 204) is estimated based on the monthly volume of deliveries that providers reported at the baseline and third intervention visits.

Consistent with other recent economic evaluations of P4P (Meacock 2014; Borghi 2015), I report cost-effectiveness results both including and excluding program start-up costs, and focus on the societal and program ICERs (excluding start-up costs) as the primary estimates for interpreting program cost-effectiveness. Although start-up costs are useful to consider as part of an *ex post* economic evaluation of the P4P initiatives, they represent sunk costs and therefore should not influence future decisions about the implementation of these P4P strategies in the region (Cellini 2010; Gold 1996). Moreover, one-time costs should ideally be annuitized based on the expected lifetime of the program (Drummond 2005), which is unknown for the Karnataka P4P interventions. By attributing all start-up costs to the first contract year, my model provides an upper limit for the start-up costs that are truly applicable to this 1-year period.

Recurrent program costs

Recurrent program costs include the costs of payouts to providers, periodic in-person visits with providers (four meetings per provider in a contract year), training of field associates who conduct the in-person provider visits, performance data collection through household interviews, and external field monitoring and support from the impact evaluation team (Appendix C.3). I calculate average performance rewards per provider based on actual payments to providers in the input and outcome P4P arms. Additional payouts for program participation and record-keeping are defined by program design.

The costs of training field associates are extracted from invoices submitted by the contractor organization that conducted the field work (Sambodhi Research & Communications Pvt. Ltd.). (Due to potentially high turnover of field associates, I categorize training costs as a yearly recurrent expenditure rather than as a start-up cost.) The unit costs of visits and performance data collection per provider are computed based on total invoiced amount for these activities, and capture the expenses associated with: field staff wages, transportation, and meals/incidentals; office supplies and printed materials for meetings; dataset compilation; administrative support; organizational overhead; and oversight of the field work by managerial staff. The unit cost of performance data collection also account for the costs associated with recruiting and training household interviewers, a different set of staff members from the field associates who conducted in-person provider visits.

Lastly, the costs of external monitoring per program year includes the annual employment costs of a full-time project manager, contracted by the impact evaluation team (COHESIVE-India; Duke University) to supplement the efforts of field managers at Sambodhi and ensure adherence to data collection protocols. This cost component is measured using original invoices from the project manager's India-based employer, and covers salary/benefits, house rent allowance, and transportation expenses.

To compute total recurrent program cost per patient delivery, the per-provider costs of all payouts and field work are divided by the expected number of deliveries per provider over a 1-year period. Training and external monitoring costs, which I treat as fixed yearly costs, are divided by the expected number of deliveries across all participating providers.

Provider effort costs

As mentioned by Meacock et al. (2014), providers can incur substantial costs as a result of their participation in P4P programs, and these costs may or may not be fully offset by the reward payouts. Under the Karnataka P4P programs, providers received an unconditional participation reward of Rs. 2,500 per study visit, an amount that likely covered the value of providers' time spent meeting with field staff.[12] It is less clear, however, whether the performance-based rewards adequately compensated providers for their quality improvement efforts. In the base-case analysis, I assume that the average performance reward is equal to the true average value of providers' increased effort, with the rationale that participating providers were aware of and accepted the price per incremental improvement listed in their contracts.

One potential threat to this assumption is that, under the prevailing payment structure (fee-for-service), Karnataka providers could conceivably have accepted the reward contracts with the intention of passing along the cost of incentivized services to the patient or insurer. As described in Appendix C.4, I explore this possibility through regression analyses of patients' self-reported total bill for their delivery at the participating provider's facility, and find no evidence of higher hospital bills within the input or outcome P4P arms, even when adjusting for differential complication risks across the arms.

Nevertheless, it is still possible that effort costs exceeded rewards if providers systematically overestimated their final payout. For example, in the input P4P contracts, the description of input performance measurement was intentionally vague to prevent "teaching to the test" (i.e., disproportionately focusing on incentivized services), and to discourage gaming attempts. The specific questions used to measure input performance were selected based on each input's observability to patients in an earlier validation study, and captured only a subset of all

[12] This conclusion is based on the estimated OB/GYN specialist salary listed in Appendix C.5, as well as the control group providers' acceptance of these participation rewards as sole compensation for engaging in the study visits.

clinically important services. Moreover, each domain score formula gave equal weight to all included inputs, regardless of any variation in the cost of delivering different inputs. Consequently, providers who focused on unmeasured or undervalued elements of input performance may have expected higher rewards than they ultimately received.

To address this possibility, I conduct an alternative analysis that varies the average cost of uncompensated provider effort from the base-case value (i.e., $0) to the maximum possible value (i.e., the maximum performance reward specified in providers' contracts minus the average performance reward that providers received).

Complication management costs

Women who experience obstetric complications are assumed to incur additional medical treatment costs beyond the normal costs of delivery, either at the original provider facility or at a higher-tier referral facility. This cost component is included under the societal perspective to capture the potential cost offsets associated with P4P. Because the per-patient costs of delivery are otherwise assumed to be common across all three model scenarios, such costs are omitted from the analyses.

The average cost of managing each complication type is extracted from secondary sources. For the base-case analysis, I use the United Nations OneHealth Costing Tool to estimate the set of resource inputs needed to manage each maternal complication according to WHO treatment protocols (IAWG-Costing 2015). The OneHealth tool follows an ingredients approach to calculate the expected cost of drugs and supplies per complication case, first estimating the average units of each item needed per case and then valuing these inputs using price listings from the Management Sciences for Health (MSH) International Drug Price Indicator Guide and the

UNICEF Supply Catalogue. The tool also provides estimates of the health care personnel time (minutes per personnel type) and length of hospital stay required per complication case; I value these additional resource requirements using country-specific unit costs per hospital bed day (which capture the "hotel" portion of hospital costs) and personnel salaries from the WHO CHOICE databases (WHO 2015d; see Appendix C.5).

In order to inform the choice of distributions for these unit costs in the probabilistic sensitivity analysis, I extract alternative complication cost estimates from published costing studies in developing countries, each of which examined complication management costs according to usual practice at the surveyed hospitals (Levin 2003; Borghi 2003; Weissman 1999). For each of the three measured complications, the base-case unit cost value derived from the OneHealth tool is within the range of literature estimates (Appendix C.5).

Currency units

The reference year for all costs is 2014. Cost estimates that are originally reported in a different year are first translated into INR using the currency exchange for that year (if not originally reported in INR), inflation-adjusted to 2014 INR using GDP deflators for India, and then translated into 2014 USD using the 2014 exchange rate of 61.02951:1.

Sensitivity analyses

Univariate and probabilistic sensitivity analyses

I conduct univariate sensitivity analyses to identify which parameters have a strong influence on model outcomes, and a probabilistic sensitivity analysis to assess the degree of certainty that an intervention is cost-effective at varying willingness-to-pay thresholds. For the

probabilistic sensitivity analysis, I use the non-parametric approach of bootstrapping to obtain 500 alternative sets of parameters from the primary trial data. Specifically, I use cluster bootstrapping to account for the correlated structure of the household survey data (Field & Welsh 2007). For each replication, I resample 135 provider clusters with replacement from the original set of participating providers, repeat the same parameter estimation steps used in the base-case analysis (e.g., fit risk equations for each maternal complication, calculate average performance payouts for each P4P strategy), and construct the hypothetical model population using the new bootstrap sample.

For other model parameters derived from secondary sources (i.e., prior publications, financial reports, or consultation with project staff), distributional assumptions are described in Table 3.3. I use a Dirichlet distribution to jointly characterize uncertainty around the proportions of maternal deaths attributable to different complication types under the status quo. Hyperparameters for this distribution are directly based on the counts of maternal deaths by primary cause from the SRS Special Survey of Deaths census report (Registrar General of India, 2006).[13] I fit beta distributions to the point estimates and uncertainty intervals (UI) given by the WHO for the MMR (India 2015 estimate: 0.00174; 80% UI: 0.00139-0.00217), and by Kassebaum et al. (2014) for the fraction of maternal deaths occurring antepartum (global 2013 estimate: 0.246; 95% UI: 0.241-0.252). I assume that uncertainty in the attributable proportion of maternal deaths for each complication is independent of uncertainty in both the overall risk of maternal death and the risks of complications under the status quo. However, case fatality rates

[13] Taking advantage of the Dirichlet-multinomial conjugate prior relationship, I use an uninformative prior of Dirichlet(α_{pph}=1, α_{pe}=1, α_{s}=1, α_{oth}=1) for the proportions of maternal deaths attributable to different complication types, which yields a posterior distribution of Dirichlet(α_{pph}=1+r_{pph}, α_{pe}=1+r_{pe}, α_{s}=1+r_{s}, α_{oth}=1+r_{oth}) given an observation of r_i maternal deaths caused by complication type *i* in the SRS Special Survey of Deaths.

for maternal complications are re-derived from these parameters within each simulation to ensure consistency between model output and all epidemiologic parameters.

With the exception of provider payouts, program cost components are assigned a Normal distribution with a coefficient of variation (ratio of standard deviation to mean) equal to 0.2 (or 0.3 for costs estimated through consultation with program staff). The values of 0.2-0.3 are arbitrary but reflect moderate uncertainty about the program costs that would be incurred in repeated iterations of the randomized experiment. Lastly, for the unit cost of managing each complication type, I assign gamma distributions with the mean equal to the point estimate from the OneHealth costing tool and standard deviation equal to the standard deviation of all alternative unit cost estimates (Appendix C.5).

In addition to the probabilistic sensitivity analysis, I perform univariate sensitivity analyses to identify parameters with a strong influence on model outcomes. The value of each parameter is varied one at a time between the lower and upper limits of its plausible range, defined as its 95% confidence interval. For each parameter examined in the univariate sensitivity analysis, I identify the threshold values at which the strategy is cost-effective at willingness-to-pay thresholds of GDP per capita and three times GDP per capita, if these threshold values fall within the parameter's plausible range.

Table 3.3: Model parameters: Base-case values and distributional assumptions

Parameter	Base-case value	Distribution in PSA	Source
Maternal mortality			
Maternal mortality ratio (maternal deaths/100,000 live births) - Status Quo	0.00174	Beta(α=34, β=19506)	WHO estimate of MMR India 2015
Fraction of maternal deaths attributable to obstetric complications - Status Quo			
Postpartum hemorrhage	0.38	Dirichlet(α_{pph}=527, α_{pe}=70, α_{s}=153, α_{oth}=637)	Registrar General, India - SRS Special Survey of Deaths - 2001-2003
Pre-eclampsia/eclampsia	0.05		
Sepsis	0.11		
Other complications and indirect causes	0.46		
Fraction of maternal deaths occurring antepartum under Status Quo	0.246	Beta(α=5,781, β=17,719)	Kassebaum 2014
Case fatality rates associated with obstetric complications			
Postpartum hemorrhage	0.001750	Bootstrap CI: 0.0012-0.0026	Derived from parameters above
Pre-eclampsia/eclampsia	0.000476	Bootstrap CI: 0.0003-0.0009	
Sepsis	0.003451	Bootstrap CI: 0.0019-0.0058	
Probability of maternal death from other causes in target population	0.000372	Bootstrap CI: 0.0003-0.0005	Derived from parameters above
Above-treatment program costs			
Start-up costs			
Cost of preparing and validating survey instruments	110,000	Normal(μ=110000, σ=33000)	Consultation with project staff
Cost of delineating program-eligible regions	25,000	Normal(μ=25000, σ=7500)	Consultation with project staff
Cost of identifying potentially eligible providers	19,497	Normal(μ=19497, σ=5849)	Consultation with project staff
Cost of screening out ineligible providers	27,871	Normal(μ=27871, σ=5574)	Invoices, Sambodhi Research & Communications Pvt Ltd
Reurrent costs			
Average performance reward per provider			
Input-based contracts	277.18	Bootstrap CI: 186-330	Primary trial data
Outcome-based contracts	954.28	Bootstrap CI: 836-1,065	Primary trial data
Participation rewards per provider	122.89	Fixed	Defined by program
Record-keeping reward per provider	16.39	Fixed	Defined by program
Cost of field staff training per contract year	4,229	Normal(μ=4,229, σ=845.8)	Invoices, Sambodhi Research & Communications Pvt Ltd
Cost per in-person visit with provider	122.24	Normal(μ=122.24, σ=24.45)	
Cost of performance data collection per provider	1,028	Normal(μ=1,028, σ=205.6)	
Cost of external project management and field monitoring per contract year	26,869	Normal(μ=26,869, σ=5,374)	Invoices, Kelly Services Pvt Ltd
Number of in-person visits per provider during contract year	4	Fixed	Defined by program
Size of target population			

Table 3.3 (Continued)

Number of program-eligible obstetric care providers in rural Karnataka	135	Fixed	Primary trial data
Number of deliveries per provider during contract year	204	Bootstrap CI: 190-226	Primary trial data
Treatment costs			
Cost of uncompensated provider effort	0	Fixed	Assumption
Average cost of complication management per case			
Postpartum hemorrhage	76.03	Gamma(α=2.609, β=29.141)	OneHealth costing tool; Levin 2003; Borghi 2003; Weissman 1999
Pre-eclampsia/eclampsia	107.95	Gamma(α=2.992, β=36.082)	
Sepsis	116.99	Gamma(α=0.959, β=122.043)	

Alternative analyses

I also conduct alternative analyses in which I vary one or more parameters based on assumptions that are either more or less conservative than those used in the base case analysis. As described earlier, one set of alternative analyses assumes that the costs of providers' quality improvement efforts are not fully offset by the financial incentives, and recalculates the societal ICER accordingly.

In another set of alternative analyses, I assume that the content and/or management of the P4P programs can be strategically modified to reduce program spending without altering their effects on maternal health outcomes. Three potential modifications are tested: (i) reducing the cost of conducting household interviews by 25%; (ii) reducing the number of in-person provider visits from four to three visits per contract year; and (iii) eliminating the need for external project management and field monitoring. Household interview costs could potentially be reduced by shortening the household survey instrument, which during the experiment took approximately one hour to administer and contained a number of items beyond those used to calculate performance. Further, it may be reasonable to eliminate one of the four in-person provider visits,

considering that the baseline visit of the experiment largely functioned as a screening visit and could be considered a one-time start-up activity rather than a yearly recurrent activity. The need for external support could be gradually reduced by integrating these project management functions into the next level of management.

RESULTS

Characteristics of patients, providers, and provider facilities

Table 3.4 summarizes the baseline characteristics of obstetric care providers and patients from the randomized experiment. Just over half of the 135 participating providers were female. The large majority (96%) had MBBS or BAMS medical credentials, and the average provider had been practicing for approximately twenty years.

Figure 3.2 reports the baseline availability of various emergency obstetric services among participating facilities, as measured through interviews with hospital staff. The capacity to deliver basic emergency obstetric services was high but not universal; for example, 85-89% of facilities regularly maintained stocks of parenteral drugs used to prevent and/or treat obstetric complications. A smaller majority of facilities (71-77%) offered caesarean section and blood transfusion services, which are key elements of comprehensive emergency obstetric care (WHO 1997).

Table 3.4: Baseline characteristics of participating obstetric care providers and surveyed patients in rural Karnataka

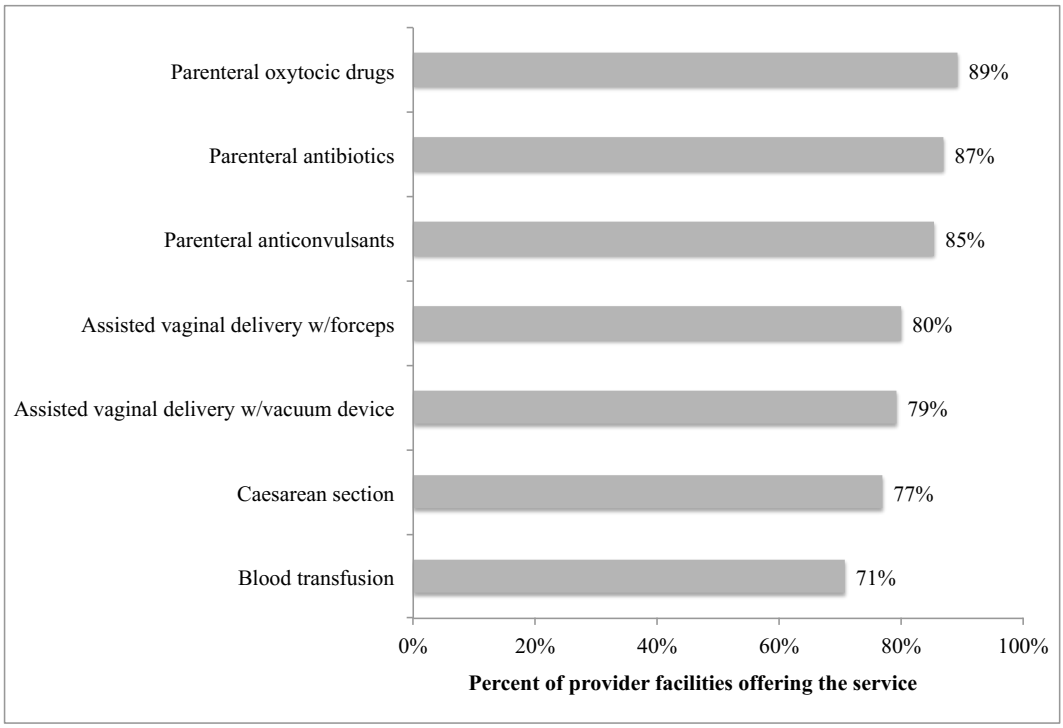
a. Provider-level characteristics

Characteristic	All providers (N=135)	Input group (N=38)	Outcome group (N=53)	Control group (N=44)
Female provider, %	55.6	55.3	56.6	54.5
Credentials, %				
MBBS	79.3	71.1	83.0	81.8
BAMS	17.0	26.3	15.1	11.4
Other qualification	3.7	2.6	1.9	6.8
Years practicing, mean (SD)	19.9 (10.7)	19.7 (10.0)	21.0 (11.0)	18.9 (11.0)

b. Patient-level characteristics

Characteristic	All surveyed households (N=2,608)	Input group (N=723)	Outcome group (N=1,078)	Control group (N=807)
Demographics				
Mother's age (years), mean (SD)	24.3 (3.8)	24.6 (4.0)	24.4 (3.8)	23.9 (3.6)
Mother's education level, %				
Illiterate	13.2	15.6	11.1	13.6
Primary only	6.1	5.4	6.5	6.3
Secondary or higher	80.7	79.0	82.4	80.0
Household's caste, %				
General or other	64.3	62.7	64.8	64.9
Scheduled caste	11.8	9.5	12.7	12.8
Scheduled tribe	7.5	6.6	7.9	7.7
Other backward class	16.4	21.2	14.6	14.6
Household owns land, %	51.7	54.1	51.5	49.9
Household has Below Poverty Line card, %	62.8	65.4	59.1	65.4
Clinical characteristics				
Mother's first pregnancy, %	49.0	47.4	52.3	45.8
Number of previous children birthed, mean (SD)	1.7 (0.9)	1.7 (0.9)	1.7 (0.9)	1.8 (1.1)
Mother has had a stillbirth or abortion, %	9.5	9.1	9.0	10.7
Comorbidity history, %				
Hypertension	6.0	9.1	5.0	4.5
Diabetes	1.0	1.1	1.3	0.6
Previous stomach surgery	11.3	10.0	10.3	13.8

Figure 3.2: Routine availability of basic and comprehensive emergency obstetric services at baseline



Notes: The routine availability of each service type was measured at the baseline or first study visit before contract implementation, based on interviews with hospital personnel. Elements of basic emergency obstetric care include the availability of parenteral drugs (i.e., those administered via injection or infusion), assisted vaginal delivery, and manual removal of placenta and of retained products (not measured). Comprehensive emergency obstetric care also includes the availability of caesarean section and blood transfusion (WHO 1997).

Base-case results

Under both the input and outcome P4P scenarios, program start-up costs are estimated at $182,368 for the region, amounting to $1,351 per participating provider (Table 3.5). Under input P4P, recurrent program costs are an estimated $292,110 during the 1-year contract period, which translates to $2,164 per participating provider or $10.61 per patient delivery. The single largest recurrent cost component (performance data generation) comprises 48% of the $292,110/year total and is 3.7 times larger than the cost of performance rewards. Recurrent program costs under the outcome P4P scenario ($383,518/year) are higher than under input P4P due to substantially higher performance rewards ($277.18 versus $954.28 per provider on average).

Table 3.5: Financial costs of P4P programs for entire study region, by program activity

a. Start-up costs

Cost component	Input P4P		Outcome P4P	
	2014 USD	%	2014 USD	%
Production and validation of survey instruments for P4P program	110,000	60	110,000	60
Identification of rural geographic areas to target	25,000	14	25,000	14
Identification of potentially eligible providers within targeted regions	19,497	11	19,497	11
Screening of ineligible providers	27,871	15	27,871	15
TOTAL START-UP COSTS	182,368	100	182,368	100

b. Recurrent costs during 1-year contract period

Cost component	Input P4P		Outcome P4P	
	2014 USD	%	2014 USD	%
Reward and participation payouts				
Participation rewards	16,590	6	16,590	4
Record-keeping rewards	2,213	1	2,213	1
Performance rewards	37,419	13	128,828	34
Provider visits & household data collection				
Training of field associates	4,229	1	4,229	1
In-person provider visits	66,010	23	66,010	17
Household data collection	138,780	48	138,780	36
External project management and field monitoring	26,869	9	26,869	7
TOTAL RECURRENT COSTS	292,110	100	383,518	100

Notes: Total program costs for the region are calculated under the hypothetical scenario that all 135 participating providers are assigned to the P4P strategy indicated.

Table 3.6a shows expected health outcomes and costs per patient delivery for each modeling scenario under base-case parameter assumptions. Incremental costs and effects (Δ) are relative to the next less expensive non-dominated scenario. Resulting ICERs (costs per life year saved and per complication averted) are presented in Table 3.6b.

Under the input P4P scenario, I estimate a reduction of 4.8 maternal complications (and 4.4 individual cases with ≥1 complications) per 100 patient deliveries compared to the status quo. This reduction in maternal complication risk is expected to result in 6.3 fewer maternal deaths, with a corresponding gain of 165 life years, per 100,000 patient deliveries. From a program perspective, the incremental cost per life year saved for the input P4P scenario is $6,418 (or $10,425 with start-up costs included), implying that the intervention is not cost-effective at a willingness-to-pay threshold of three times GDP per capita ($4,744.50; World Bank 2015). When taking a societal perspective, program costs associated with input P4P are partially offset by reductions in the direct medical costs of managing complications, with an estimated $1.91 saved on medical treatment per woman delivering during the contract period. Nevertheless, base-case ICERs from the societal perspective ($5,262 and $9,269 excluding or including start-up costs, respectively) are above the three times GDP per capita threshold. With input P4P, the cost per complication averted ranges from $180 to $356 depending on the perspective and scope of costs.

As shown, outcome P4P is a dominated strategy regardless of perspective, with higher costs and lower health benefits than both the status quo and input P4P scenarios.

Table 3.6: Base-case estimates of effects, costs, and incremental cost-effectiveness

a. Expected health outcomes and costs per patient in the target population

Outcome	Status Quo	Input P4P		Outcome P4P	
	Estimate	Estimate	Δ vs. Status Quo	Estimate	Δ vs. Input P4P
Clinical outcomes of delivery					
Risk of obstetric complications					
Postpartum hemorrhage	0.378	0.282	-0.096	0.367	0.085
Pre-eclampsia/eclampsia	0.183	0.203	0.020	0.221	0.018
Sepsis	0.055	0.083	0.028	0.087	0.004
Any complication (≥1 of the above)	0.510	0.466	-0.044	0.539	0.073
Number of complications per patient	0.616	0.568	-0.048	0.676	0.108
Mortality and life expectancy					
Risk of maternal death					
Overall	0.001312	0.001249	-0.000063	0.001422	0.000173
By cause					
Postpartum hemorrhage	0.000661	0.000493	-0.000168	0.000642	0.000149
Pre-eclampsia/eclampsia	0.000087	0.000097	0.000010	0.000105	0.000009
Sepsis	0.000191	0.000286	0.000095	0.000302	0.000015
Other causes	0.000372	0.000372	0.000000	0.000372	0.000000
Expected life-years (with 3% annual discounting)	26.249534	26.251187	0.001653	26.246640	-0.004547
Costs per patient (2014 USD)					
Start-up costs	0.00	6.62	6.62	6.62	0.00
Recurrent program costs	0.00	10.61	10.61	13.93	3.32
Direct medical costs of complication management	54.97	53.06	-1.91	62.04	8.99
Uncompensated provider effort costs	0.00	0.00	0.00	0.00	0.00

Target population includes women who present to participating providers for labor/delivery within the contract year.

b. Incremental cost-effectiveness ratios by perspective and scope of costs

ICER	Status Quo	Input P4P	Outcome P4P
Cost per life-year saved			
Societal perspective (excl. start-up costs)	-	5,262	Dominated
Societal perspective (incl. start-up costs)	-	9,269	Dominated
Program perspective (excl. start-up costs)	-	6,418	Dominated
Program perspective (incl. start-up costs)	-	10,425	Dominated
Cost per complication averted			
Societal perspective (excl. start-up costs)	-	180	Dominated
Societal perspective (incl. start-up costs)	-	317	Dominated
Program perspective (excl. start-up costs)	-	219	Dominated
Program perspective (incl. start-up costs)	-	356	Dominated

Univariate and probabilistic sensitivity analysis results

Table 3.7 presents findings from the univariate and threshold sensitivity analyses for both the societal perspective (Table 3.7a) and program perspective (Table 3.7b). To highlight variables with a strong influence on the cost-effectiveness of input P4P, parameters are sorted from widest to narrowest range of ICER values over the plausible range of parameter values. (Because outcome P4P is dominated in the base-case analysis and nearly all probabilistic simulations, I focus exclusively on input P4P for the univariate sensitivity analyses.)

The ICER for input P4P is highly sensitive to the intervention's effectiveness in preventing postpartum hemorrhage relative to the status quo. Holding all else constant, input P4P is estimated to be cost-effective at a willingness-to-pay threshold of three times GDP per capita if the odds ratio of postpartum hemorrhage is at least as low as 0.601 or 0.568 (depending on perspective); these threshold parameter values are only slightly lower than the base-case odds ratio of 0.610. However, input P4P is a dominated strategy when varying this odds ratio to the high end of its plausible range (0.991). From the societal perspective, the cost-effectiveness of input P4P is also sensitive to the unit cost of managing postpartum hemorrhage and, to a lesser extent, the unit costs of managing other complications. (Complication management costs do not impact cost-effectiveness from the program perspective.)

Other variables that are reasonably influential include the maternal mortality ratio (which proportionately affects the estimated mortality burden of each maternal complication), and the average volume of deliveries per provider during the contract year (due to economies of scale when assuming a higher number of deliveries per provider). However, from both the societal and program perspectives, input P4P remains cost-ineffective at the three times GDP per capita threshold across the plausible range of average performance rewards per provider.

Table 3.7: Univariate and threshold sensitivity analyses

a. Societal perspective (excluding start-up costs)

Parameter	Range of ICERs: Input P4P vs. Status Quo (2014 USD per LY saved)			Plausible Range of Parameter Values[a,b]				
	Base-case	Low	High	At base-case ICER	At low ICER	At high ICER	At GDP per capita	At 3*GDP per capita
Odds ratio for postpartum hemorrhage[c]	5,262	50	Dom	0.610	0.301	0.991	0.481	0.601
Unit cost of managing postpartum hemorrhage	5,262	40	8,921	76	166	13	140	85
Unit cost of managing sepsis	5,262	3,364	9,242	117	3	356	n/a	86
Maternal mortality ratio[d]	5,262	3,864	7,568	0.00174	0.00237	0.00121	n/a	0.00193
Unit cost of managing pre-eclampsia	5,262	4,219	6,707	108	22	227	n/a	65
Deliveries per provider in contract year	5,262	4,638	5,735	204	226	190	n/a	222
Average performance rewards per provider	5,262	4,992	5,419	277	186	330	n/a	n/a

b. Program perspective (excluding start-up costs)

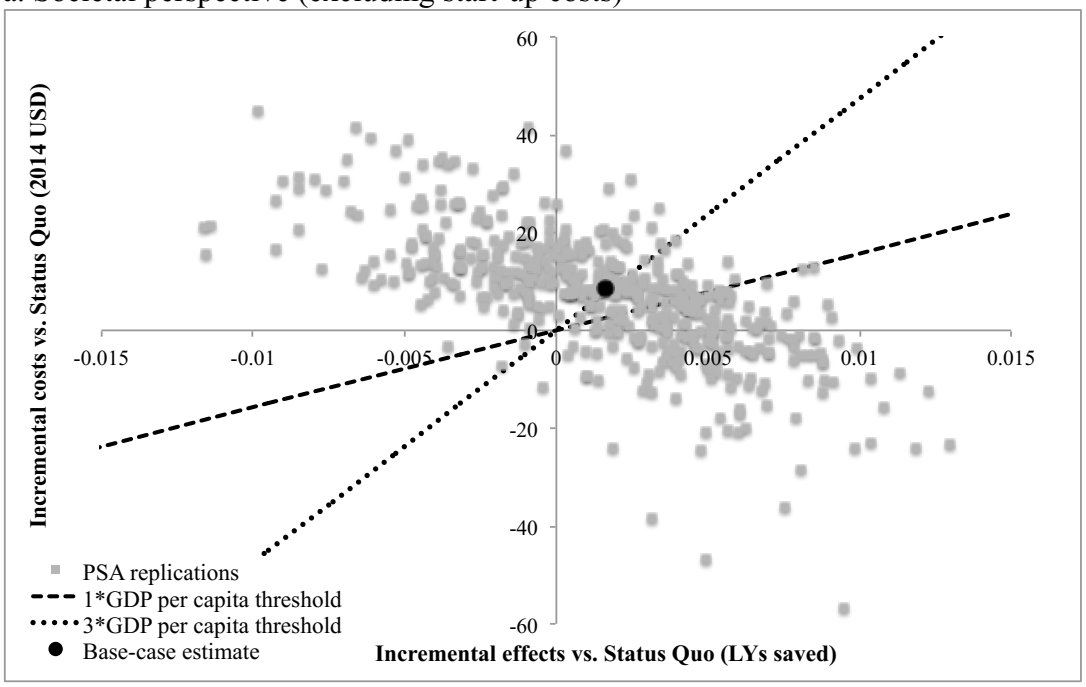
Parameter	Range of ICERs: Input P4P vs. Status Quo (2014 USD per LY saved)			Plausible Range of Parameter Values[a,b]				
	Base-case	Low	High	At base-case ICER	At low ICER	At high ICER	At GDP per capita	At 3*GDP per capita
Odds ratio for postpartum hemorrhage[c]	6,418	1,582	Dom	0.610	0.301	0.991	0.301	0.568
Maternal mortality ratio[d]	6,418	4,712	9,229	0.00174	0.00237	0.00121	n/a	0.00235
Deliveries per provider in contract year	6,418	5,793	6,891	204	226	190	n/a	n/a
Average performance rewards per provider	6,418	6,147	6,574	277	186	330	n/a	n/a

[a] The plausible range for each parameter corresponds to the 95% confidence interval (based on the empirical bootstrap distribution or the parametric distribution specified in Table 3.3).
[b] A parameter value of 'n/a' at the GDP per capita ($1,582/LY) or three times GDP per capita ($4,745/LY) willingness-to-pay threshold indicates that the ICER for input P4P does not fall below this threshold across the range of plausible values for that parameter.
[c] For the univariate sensitivity analysis of the postpartum hemorrhage odds ratio (input P4P vs. status quo), I vary the odds ratio value and otherwise leave the risk equation for postpartum hemorrhage unchanged.
[d] The case fatality rates for all three measured maternal complications change proportionately when the maternal mortality ratio (MMR) is varied.

Results from the probabilistic sensitivity analysis are summarized through scatter plots of simulated incremental cost and effect pairs (Figure 3.3), and cost-effectiveness acceptability curves indicating the probability that each policy scenario is cost-effective at various willingness-to-pay thresholds (Figure 3.4). As illustrated by the scatterplots, the incremental cost of input P4P is subject to far greater uncertainty under the societal perspective than under the program perspective. This result is due to uncertainty surrounding the unit costs of complication management, given the wide range of literature estimates for these parameters. Nevertheless, at a willingness-to-pay threshold of three times GDP per capita, the probability that input P4P is cost-effective is similar between the societal and program perspectives (49% and 45%, respectively) when focusing on recurrent costs only (Figure 3.4a-b). When also including start-up costs, input P4P is cost-effective in 33% and 42% of simulations under the societal and program perspectives, respectively (Figure 3.4c-d).

Figure 3.3: Scatterplot: Joint density of incremental costs and effects under input P4P from the probabilistic sensitivity analysis

a. Societal perspective (excluding start-up costs)



b. Program perspective (excluding start-up costs)

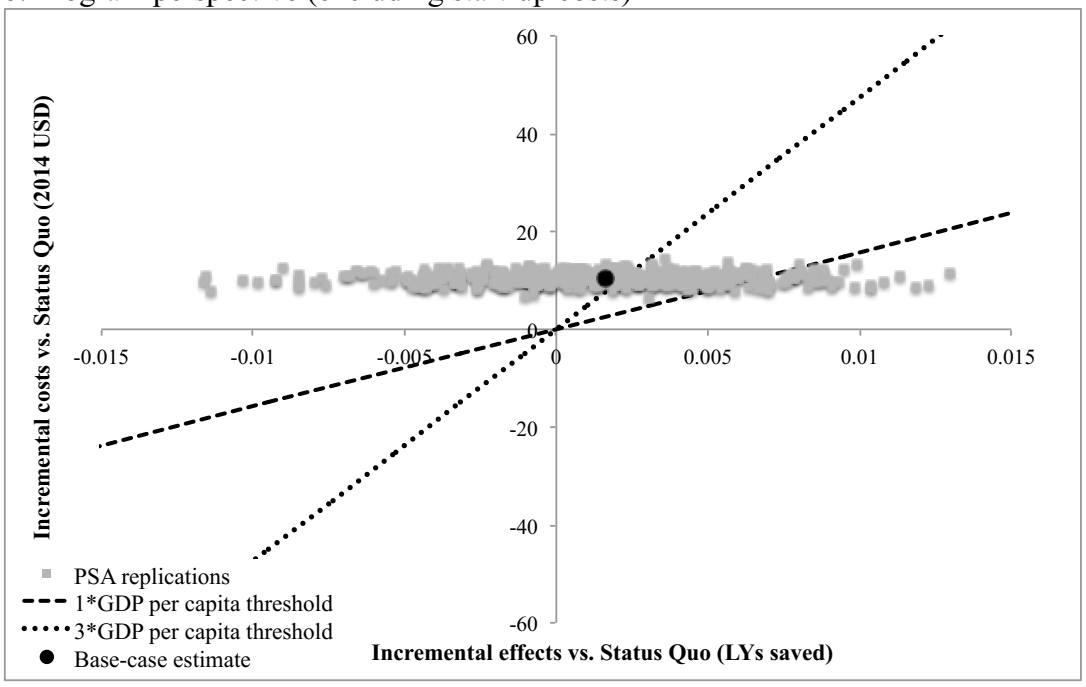
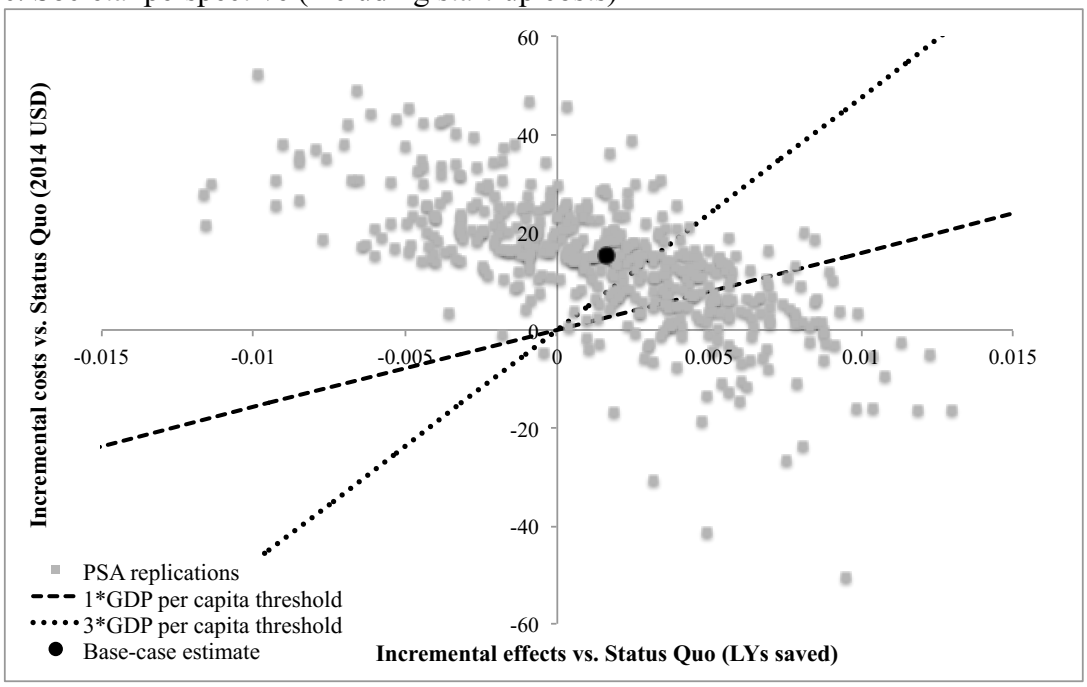
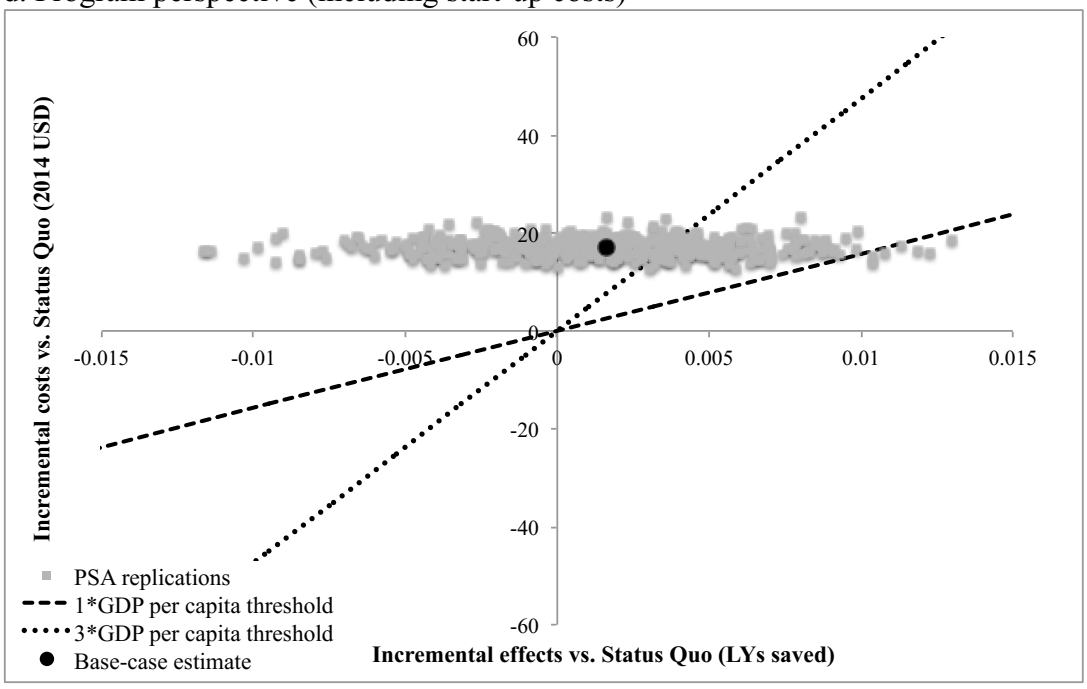


Figure 3.3 (Continued)

c. Societal perspective (including start-up costs)



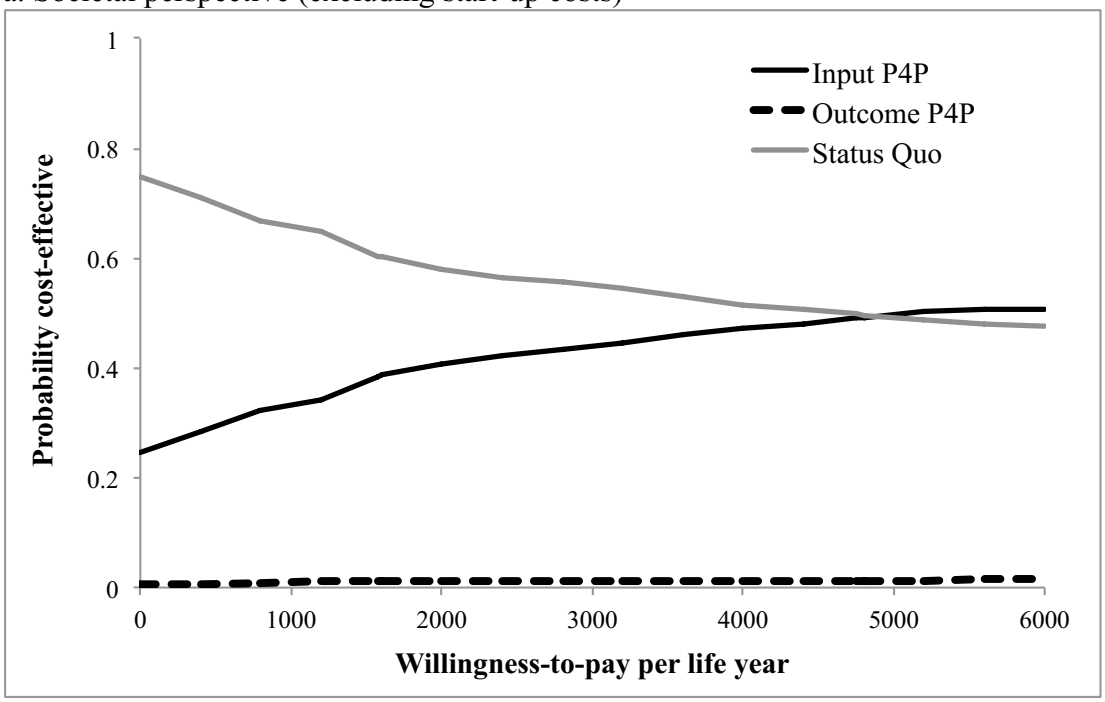
d. Program perspective (including start-up costs)



Notes: Because outcome P4P is nearly always a dominated strategy in the probabilistic sensitivity analysis, scatterplots focus on incremental costs and effects for input P4P versus the status quo. Incremental cost and effect pairs are cost-effective at the 1* or 3*GDP per capita threshold if they fall below the corresponding threshold line.

Figure 3.4: Cost-effectiveness acceptability curves

a. Societal perspective (excluding start-up costs)



b. Program perspective (excluding start-up costs)

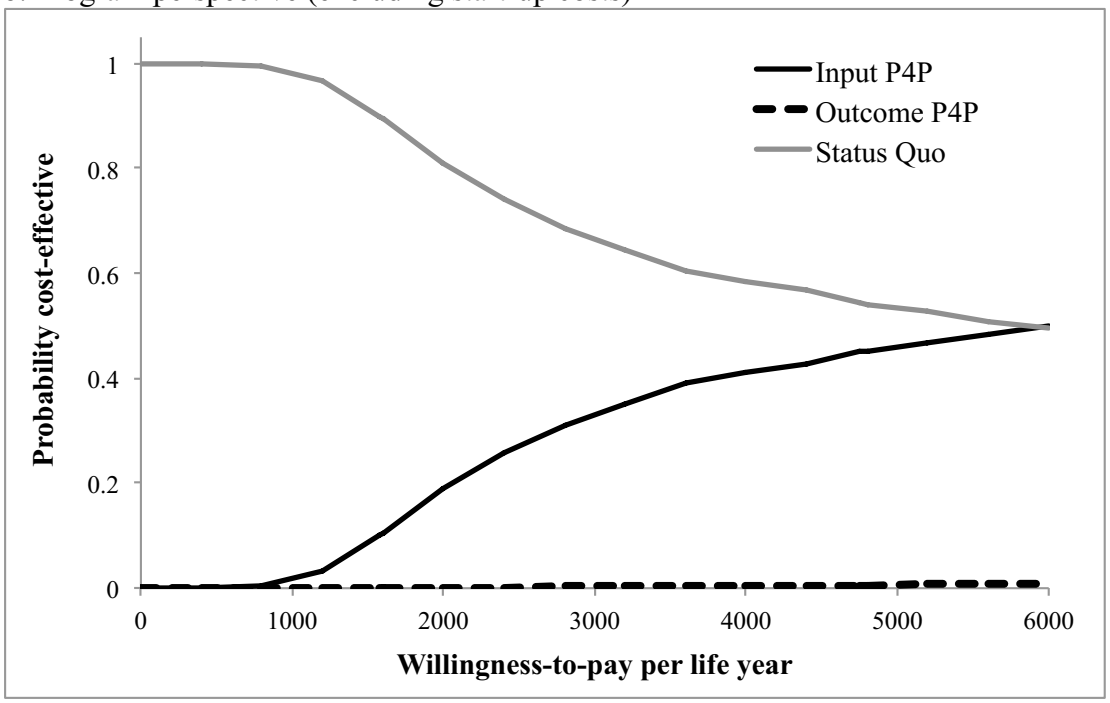
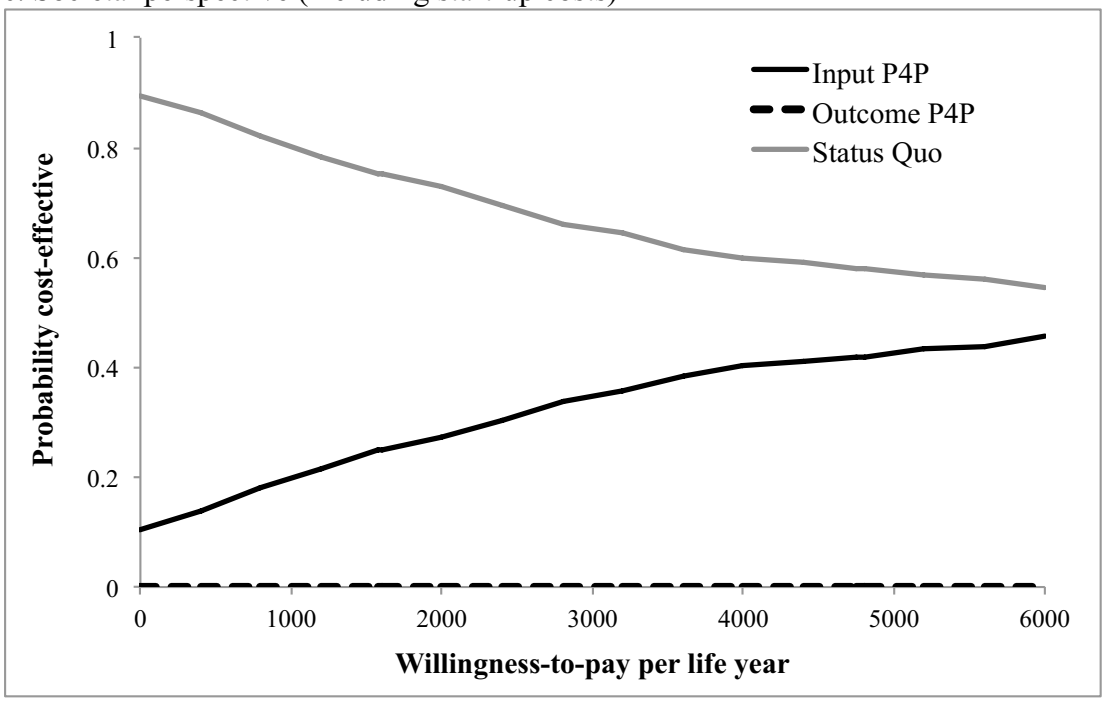
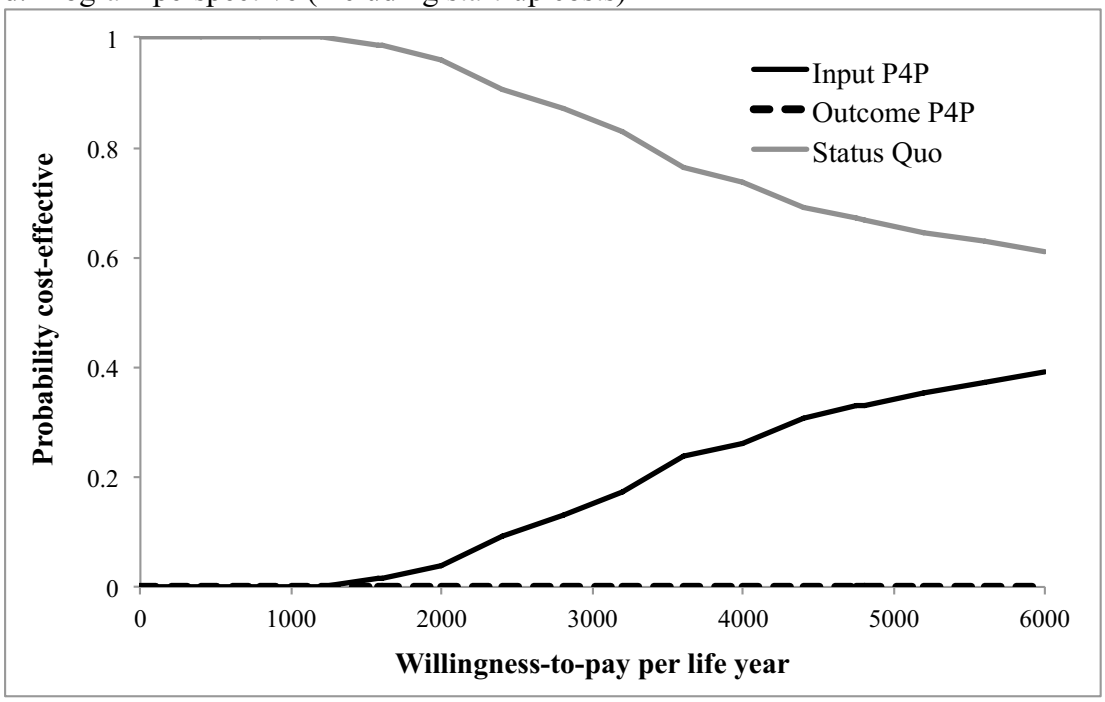


Figure 3.4 (Continued)

c. Societal perspective (including start-up costs)



d. Program perspective (including start-up costs)



Alternative analysis results

The base-case and sensitivity analyses above make the simplifying assumption that, on average, the total economic cost of providers' quality improvement effort is equal to performance reward payouts. To the extent that there are uncompensated provider effort costs, the input P4P strategy becomes less economically attractive from a societal perspective (Table 3.8). The contracts for input P4P explicitly state that providers can earn up to a maximum of Rs. 169,750 ($2,781), an amount that is approximately ten times as large as average performance payouts under this form of P4P ($277.18). Thus, in Table 3.8, 900% of the average performance reward represents the theoretical upper limit for the average cost of uncompensated provider effort (assuming that providers do not pass along the cost of quality improvements to patients).

Under the most conservative assumption that all providers expect to earn the maximum and actually expend this value of effort (e.g., through increases in personnel time and medical supply usage), the societal cost per life year more than doubles from $5,262 (base-case) to $12,662. Conversely, if uncompensated effort cost is similar or smaller in magnitude compared to reward payouts (≤100% of rewards), the societal ICER for input P4P increases by under 16%.

Table 3.8: Alternative analysis: Cost-effectiveness of input P4P under varying levels of uncompensated provider effort

Average cost of uncompensated provider effort, as a % of average performance rewards under input P4P (277.18 USD)	Incremental costs and effects (Input P4P vs. Status Quo)			ICER (2014 USD per life year saved)	
	ΔRecurrent program costs	ΔTreatment costs	ΔLife years	Societal perspective	Program perspective
0% (base-case)	10.61	-1.91	0.001653	5,262	6,418
20% (55.44 USD)	10.61	-1.64	0.001653	5,427	6,418
50% (138.59 USD)	10.61	-1.23	0.001653	5,674	6,418
100% (277.18 USD)	10.61	-0.55	0.001653	6,085	6,418
300% (831.54 USD)	10.61	2.17	0.001653	7,729	6,418
900% (2,494.62 USD)	10.61	10.32	0.001653	12,662	6,418

Notes: Uncompensated provider effort cost refers to the total economic cost resulting from a provider's quality improvement effort minus her compensation from performance rewards. As described in the main text, the theoretical maximum of average uncompensated effort cost is approximately equal to 900% of the $277.18 average performance payout. The cost of uncompensated effort is categorized as a treatment cost (as opposed to an above-treatment program cost), and therefore does not impact cost-effectiveness from the program perspective.

Table 3.9 shows how the cost-effectiveness of input P4P might change if program funders can successfully reduce program resource use without adversely affecting maternal health outcomes. In the event that all three potential program cutbacks are implemented, the societal and program ICERs for input P4P are both estimated to decrease by $1,715, and fall below the three times GDP per capita willingness-to-pay threshold. The first cutback alone (reducing costs of performance data collection by 25%) is estimated to reduce both ICERs by $762, placing the societal ICER for input P4P just below the three times GDP per capita threshold.

Table 3.9: Alternative analysis: Cost-effectiveness of input P4P following potential improvements in program efficiency

Potential program cost reduction	Incremental costs and effects (Input P4P vs. Status Quo)			ICER (2014 USD per life year saved)	
	ΔRecurrent program costs	ΔTreatment costs	ΔLife years	Societal perspective	Program perspective
No program cutbacks (base-case)	10.61	-1.91	0.001653	5,262	6,418
(1) Reduce household interview costs by 25%	9.35	-1.91	0.001653	4,500	5,656
(2) Reduce number of in-person provider visits from 4 to 3 visits per contract year, plus (1)	8.75	-1.91	0.001653	4,138	5,293
(3) No external project management or support, plus (1) & (2)	7.77	-1.91	0.001653	3,547	4,703

DISCUSSION

Key findings and policy implications

P4P initiatives have become increasingly common in low- and middle-income countries (Witter 2013), but there is little rigorous evidence on the costs or cost-effectiveness of such programs. Using data from a cluster-randomized trial conducted in rural Karnataka, India, I evaluate the cost-effectiveness of two P4P interventions that aim to improve the quality of maternal health care by linking financial incentives to either health service inputs or health outcomes.

Results suggest that the input-based P4P program as currently designed may not be a cost-effective strategy for reducing maternal mortality in the region. Under base-case assumptions, the societal ICER for input P4P versus the status quo ($5,262 per life year saved) slightly exceeds the common willingness-to-pay threshold of three times GDP per capita

($4,745). This result is dependent on a number of factors, including assumptions about intervention effectiveness, the direct medical costs of managing complications, and the mortality burden of complications. In the probabilistic sensitivity analysis, input P4P is cost-effective in just under half of the simulations when focusing on recurrent costs only.

During the experiment, providers assigned to input P4P achieved reductions in the occurrence of postpartum hemorrhage, the leading cause of maternal deaths in India (Registrar General, India 2006). However, the lack of improvement on other major obstetric complications limits the cost-effectiveness of this incentive scheme. Providers may have intentionally prioritized the prevention of postpartum hemorrhage based on perceived need: At baseline, 74% of providers reported hemorrhage as the most clinically important outcome to improve (Mohanan et al.; forthcoming). Alternatively, the risk of postpartum hemorrhage may have simply been more amenable to change through guideline adherence than other complications, given that effective primary prevention measures are available for hemorrhage; for example, while timely diagnosis and treatment can reduce the risk of progression from pre-eclampsia to eclampsia, there is little guidance available on how to prevent pre-eclampsia (WHO 2011).

Outcome-based P4P is a strongly dominated strategy in the base-case analysis and in nearly all probabilistic simulations, as expected given that this strategy did not show health benefits in the trial. Interestingly, despite worse maternal health outcomes under outcome P4P, expected performance rewards per provider are considerably higher under this strategy compared to input P4P. This paradox appears to stem from a usual limitation of outcome-based performance measures; namely, such measures give a "noisy" signal of performance because adverse outcomes do not always occur when there are quality deficits (Petersen 2006). Due to stochastic variability in the outcomes of each provider's patients, and the fact that rewards can

only be zero or positive, average performance rewards are substantial under outcome P4P – even though average complication risks under outcome P4P are slightly worse than the pre-specified performance benchmarks (the $\overline{y}_i$'s in the Methods section). In addition to causing high program costs, the noisiness of outcome-based performance measures might also explain the ineffectiveness of this scheme: If providers are risk-averse and view outcomes as largely beyond their control, they may be reluctant to undertake quality improvements out of concern that their efforts will go unrewarded (Miller 2013).

Under input P4P, the main driver of program spending is by far the collection of performance data, reaffirming earlier arguments that cost-effectiveness analyses of P4P should consider a broad scope of program costs beyond just the performance-based incentives (Emmert 2012; Meacock 2014). Conducting household interviews for the randomized experiment was a difficult undertaking that involved over thirty interviewers and seven field associates and managers who supervised the interviewer staff, plus office personnel and an additional level of oversight from an external project manager. Based on documented expenditures during the experiment, performance data collection comprises an estimated 48% of recurrent program costs and outweighs input-based performance rewards by a factor of 3.7. In Borghi et al. (2015), a study that evaluated the cost-effectiveness of a P4P pilot program in Tanzania, data collection accounted for a similarly large proportion of program running costs (36-51% of annual program costs, depending on the scenario being modeled). These results highlight a key barrier to implementing efficient P4P programs in resource-limited settings without routine health information systems already in place. Data collection activities may be a less important cost driver in settings that, for example, supported web-based patient reported outcomes.

Although input P4P is most likely not a cost-effective program by its current design, it may be possible to improve the program's efficiency by fine-tuning its features and management structure. Scenario analyses indicate that the input P4P program is more likely to be cost-effective if the cost of data collection can be reduced by one-quarter. Such cost savings might be achievable by shortening the household survey instrument, which would lessen the personnel requirements for conducting interviews, or by dedicating governmental staff within each district to conduct household interviews in order to reduce year-to-year turnover of interviewers and reduce travel expenditures. The cost-effectiveness of input P4P would further improve if, in addition to reducing data collection costs, the Government of Karnataka is able to gradually reduce the need for external field monitoring and support. Reducing the number of in-person provider visits from four to three per contract year would also help reduce program costs, although it would be important to monitor the effects of such a change on provider performance.

To date, P4P initiatives have predominantly used process-of-care indicators instead of health outcomes to measure performance, which could explain the limited consideration of cost-effectiveness in the P4P literature. Under input-based P4P arrangements, the performance data collection may not yield suitably proximate measures of clinical benefit that can be linked to typical measures of health benefit in cost-effectiveness analyses (e.g., life-years, quality-adjusted life-years). Therefore, one key strength of the present analysis is the availability of clinically meaningful outcome measures that are close antecedents of maternal mortality, enabling me to conduct a full economic evaluation even for the input P4P strategy. To my knowledge, this study is only the second cost-effectiveness analysis of P4P in a resource-limited setting (Borghi et al. 2015), and the first of such studies to report results using a standard metric (i.e., cost per life year saved).

Limitations

This study is subject to several limitations, most of which would tend to create a bias against performance-based incentives. First, intervention effectiveness parameters in the model are based on comparisons against the control group in the randomized experiment, which assumes that control group outcomes are reasonably representative of the status quo. The denominator of the ICER for input P4P therefore nets out any incremental benefit that may have resulted from the in-person provider visits, provision of educational materials, and participation rewards alone, even though the cost of these ancillary program features is included in the numerator.

Second, the model assumes that any quality improvements under P4P are temporary and only affect the maternal health outcomes of patients who present for delivery during the contract year. If input-based P4P produces sustained reductions in postpartum hemorrhage among future patients who deliver after the contract period ends, the cost-effectiveness of this strategy could be underestimated.

Third, I assume no potential for benefit from performance-based incentives beyond the measured difference in maternal complication risks. For example, the model does not consider the potential impact of P4P interventions on the risk of death conditional on experiencing a particular complication, which could occur if P4P led to more effective management of complications (or more timely referrals to higher-tier facilities, if a provider lacks the capacity to deliver comprehensive emergency obstetric care). Further, as noted earlier, trial data suggest that input P4P likely encouraged the delivery of uterine massage and administration of oxytocic drugs. In addition to preventing postpartum hemorrhage, such clinical actions could conceivably

change the composition and severity of hemorrhage cases that do occur, another potential mechanism by which financial incentives might affect the case fatality rate.

Fourth, provider rewards are taken as an approximation of the true economic costs of the providers' increased efforts. To address this limitation, I present sensitivity analyses that assume varying levels of uncompensated provider effort costs. Note that while uncompensated provider effort does not affect costs from the program perspective during the first contract period, this cost burden could eventually shift to program funders if they wish to renew contracts with the same providers. To the extent that providers received low compensation relative to effort in previous contract periods, funders may need to increase the generosity of rewards to maintain the same level of engagement.

Lastly, as with any cost-effectiveness analysis, results are subject to uncertainty about the true value of model parameters. Sensitivity analyses show that the ICER for input P4P is highly sensitive to the effectiveness of this strategy in reducing the risk of postpartum hemorrhage. Although the trial-based odds ratio of postpartum hemorrhage with input P4P versus controls is statistically significant, the confidence interval around this point estimate is wide due to sample size limitations. Updated cost-effectiveness analyses are warranted if, through further experimentation, new data become available on the effectiveness of this P4P strategy to prevent maternal complications.

Conclusions

During a randomized experiment in rural Karnataka, India, providers who were rewarded for adherence to obstetric care guidelines achieved fewer maternal complications compared to providers who were offered no performance-based incentives. However, the costs of generating

performance data and engaging with providers were substantial in this resource-limited setting. Given a willingness-to-pay threshold of three times GDP per capita, the input-based P4P program as currently designed may not be a cost-effective strategy for improving the quality of maternal health care. In future trials or implementation efforts, it may be possible to improve the cost-effectiveness of this strategy if program activities can be adjusted to decrease operating costs while maintaining similar quality improvement effects.

REFERENCES

ACOG Committee on Practice Bulletins. ACOG Practice Bulletin: Clinical management guidelines for obstetrician-gynecologists: cervical cytology screening. Obstet Gynecol 2003;102:417-27.

ACOG Committee on Practice Bulletins-Gynecology. ACOG Practice Bulletin no. 109: Cervical cytology screening. Obstet Gynecol 2009;114:1409-20.

ACOG Committee on Practice Bulletins-Gynecology. ACOG Practice Bulletin no. 131: Screening for cervical cancer. Obstet Gynecol 2012;120:1222-38.

Anhang PR, Tiro JA, Saraiya M, et al. Use of human papillomavirus vaccines among young adult women in the United States: an analysis of the 2008 National Health Interview Survey. Cancer 2011;117:5560-8.

Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011;46:399-424.

Bahmanyar ER, Paavonen J, Naud P, et al. Prevalence and risk factors for cervical HPV infection and abnormalities in young adult women at enrolment in the multinational PATRICIA trial. Gynecol Oncol 2012;127:440-50.

Barter D, Brower CH, Nigam A, et al. Policy instruments to improve intervention uptake and provider quality. Disease Control Priorities in Developing Countries, 3rd edition. Working paper. Available at <http://dcp-3.org/sites/default/files/resources/Uptake%20%26%20Quality%20WP8.pdf>.

Bauer HM, Wright G, Chow J. Evidence of human papillomavirus vaccine effectiveness in reducing genital warts: an analysis of California public family planning administrative claims data, 2007-2010. Am J Public Health 2012;102:833-5.

Borghi J, Hason K, Adjei Acquah C, et al. Costs of near-miss obstetric complications for women and their families in Benin and Ghana. Health Policy Plan 2003;18:383-390.

Borghi J, Little R, Binyaruka P, et al. In Tanzania, the many costs of pay-for-performance leave open to debate whether the strategy is cost-effective. Health Aff 2015;34;406-14.

Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytical strategies using propensity scores. Ann Intern Med 2002;137:693-5.

Brewer NT, Chung JK, Baker HM, et al. Pharmacist authority to provide HPV vaccine: novel partners in cervical cancer prevention. Gynecol Oncol 2014;132:S3-8.

Brown DR, Kjaer SK, Sigurdsson K, et al. The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due

to oncogenic nonvaccine HPV types in generally HPV-naive women aged 16–26 years. J Infect Dis 2009;199:926-35.

Busch SH, Barry CL, Vegso SJ, et al. Effects of a cost-sharing exemption on use of preventive services at one large employer. Health Aff 2006;25:1529-36.

Bynum SA, Staras SA, Malo TL, Giuliano AR, Shenkman E, Vadaparampil ST. Factors associated with Medicaid providers' recommendation of the HPV vaccine to low-income adolescent girls. J Adolesc Health 2014;54:190-6.

Cellini SR, Kee JE. "Cost-effectiveness and cost-benefit analysis," in Wholey JS, Hatry HP, Newcomer KE, eds., Handbook of Practical Program Evaluation, 3rd edition. 2010. San Francisco, CA: Jossey-Bass.

Centers for Disease Control and Prevention. Recommendations on the use of quadrivalent human papillomavirus vaccine in males--Advisory Committee on Immunization Practices (ACIP), 2011. MMWR 2011;60:1705-8.

Centers for Disease Control and Prevention. Recommended and minimum ages and intervals between doses. 2012. Available at <http://www.cdc.gov/vaccines/pubs/pinkbook/downloads/appendices/a/age-interval-table.pdf>.

Cherkin DC, Grothaus L, Wagner EH. The effect of office visit copayments on utilization in a health maintenance organization. Med Care 1990;27:1036-45.

Chernew M, Gibson TB, Yu-Isenberg K, et al. Effects of increased patient cost sharing on socioeconomic disparities in health care. J Gen Intern Med 2008;23:1131-6.

Chernew ME, Shah MR, Wegh A, et al. Impact of decreasing copayments on medication adherence within a disease management environment. Health Aff 2008;27:103-12.

Chesson HW, Ekwueme DU, Saraiya M, et al. Cost-effectiveness of human papillomavirus vaccination in the United States. Emerg Infect Dis 2008;14:244-51.

Choudhry NK, Fischer MA, Avorn J, et al. At Pitney Bowes, value-based insurance design cut copayments and increased drug adherence. Health Aff 2010;29:1995-2001.

Crowe E, Pandeya N, Brotherton JML, et al. Effectiveness of quadrivalent human papillomavirus vaccine for the prevention of cervical abnormalities: case-control study nested within a population based screening programme in Australia. BMJ 2014;348:1458.

Das J, Hammer J. Are institutional births institutionalizing deaths? World Bank Future Development blog, 2014 Nov 20. Available at <http://blogs.worldbank.org/futuredevelopment/are-institutional-births-institutionalizing-deaths>.

Das J, Holla A, Das V, et al. In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps. Health Aff 2012;31:2774-84.

Das J, Hammer J, Sanchez-Paramo C. The impact of recall periods on reported morbidity and health seeking behavior. J Dev Econ 2012b;98:76-88.

Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. J Econ Perspect 2008;22:93-114.

Department of Health and Human Services. Women's preventive services guidelines. 2012 Aug 1. Available at < http://www.hrsa.gov/womensguidelines/>.

Department of the Treasury, Department of Labor, Department of Health and Human Services. Interim final rules for group health plans and health insurance coverage relating to status as a grandfathered health plan under the Patient Protection and Affordable Care Act. Interim final rules with request for comments. Fed Regist 2010 Jun 17;75:34537-70.

Department of the Treasury, Department of Labor, Department of Health and Human Services. Interim final rules for group health plans and health insurance insurers relating to coverage of preventive services under the Patient Protection and Affordable Care Act. Federal Registrer. 2010 July 19;75:41726-60.

de Bruin SR, Baan CA, Struijs JN. Pay-for-performance in disease management: a systematic review of the literature. BMC Health Serv Res 2011;11:272.

de Sanjose S, Quint WG, Alemany L, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol 2010;11:1048-56.

Dobson SR, McNeil S, Dionne M, et al. Immunogenicity of 2 doses of HPV vaccine in younger adolescents vs 3 doses in young women: a randomized clinical trial. JAMA 2013;309:1793-802.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes: 3rd ed. 2005. Oxford University Press: New York.

Dunne EF, Markowitz LE, Saraiya M, et al. CDC grand rounds: reducing the burden of HPV-associated cancer and disease. MMWR 2014;63:69-72.

Elam-Evans LD, Yankey D, Jeyarajah J, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13–17 years--United States, 2013. MMWR Morb Mortal Wkly Rep 2014;63:625-33.

Eijkenaar F. Pay for performance in health care: an international overview of initiatives. Med Care Res Rev 2012;69:251-76.

Eijkenaar F, Emmert M, Scheppach M, et al. Effects of pay for performance in health care: a systematic review of systematic reviews. Health Policy 2013;110:115-30.

Elbasha EH, Dasbach EJ, Insinga RP. Model for assessing human papillomavirus vaccination strategies. Emerg Infect Dis 2007;13:28-41.

Emmert M, Eijkenaar F, Kemter H, et al. Economic evaluation of pay-for-performance in health care: a systematic review. Eur J Health Econ 2012;13:755-67.

Fagot JP, Boutrelle A, Ricordeau P, et al. HPV vaccination in France: uptake, costs and issues for the National Health Insurance. Vaccine 2011;29:3610-6.

Fendrick AM, Martin JJ, Weiss AE. Value-based insurance design: more health at any price. Health Serv Res 2012;47:404-13.

Field CA, Welsh AH. Bootstrapping clustered data. J R Statist Soc 2007;69:369-90.

Filippi V, Ronsmans C, Gandaho T, et al. Women's reports of severe (near-miss) obstetric complications in Benin. Stud Fam Plann 2000;31:309-24.

Ford CA, English A, Davenport AF, et al. Increasing adolescent vaccination: barriers and strategies in the context of policy, legal, and financial issues. J Adolesc Health 2009;44:568-74.

Francis SA, Nelson J, Liverpool J, et al. Examining attitudes and knowledge about HPV and cervical cancer risk among female clinic attendees in Johannesburg, South Africa. Vaccine 2010;28:8026-32.

Frank MB, Fendrick AM, He Y, et al. The effect of a large regional health plan’s value-based insurance design program on statin use. Med Care 2012;50:934-9.

FUTURE I/II Study Group, Dillner J, Kjaer SK, et al. Four year efficacy of prophylactic human papillomavirus quadrivalent vaccine against low grade cervical, vulvar, and vaginal intraepithelial neoplasia and anogenital warts: randomised controlled trial. BMJ 2010;341:c3493.

Gertig DM, Brotherton JML, Budd AC, et al. Impact of a population-based HPV vaccination program on cervical abnormalities: a data linkage study. BMC Medicine 2013;11:227.

Gilkey MB, Malo TL, Shah PD, et al. Quality of physician communication about human papillomavirus vaccine: findings from a national survey. Cancer Epidemiol Biomarkers Prev 2015;24;1-7.

Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. Oxford University Press: New York. 2006.

Goldie SJ, Sweet S, Carvalho N, et al. Alternative strategies to reduce maternal mortality in India: a cost-effectiveness analysis. PLoS Med 2010;7:e1000264.

Greenland S. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. Epidemiology 2003;14:300-306.

Hamel L, Firth J, Brodie M. Kaiser Health Tracking Poll: March 2014. Available at <http://kff.org/health-reform/poll-finding/kaiser-health-tracking-poll-march-2014/>.

Hariri S, Unger ER, Powell SE, et al. Human papillomavirus genotypes in high-grade cervical lesions in the United States. J Infect Dis 2012;206:1878-86.

Healthy People 2020: Immunization and infectious diseases. Available at <https://www.healthypeople.gov/2020/topics-objectives/topic/immunization-and-infectious-diseases/objectives>.

Henrikson NB, Anderson ML, Hubbard RA, Fishman P, Grossman DC. Employee knowledge of value-based insurance design benefits. Am J Prev Med 2014;47:115-22.

Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561-70.

Hernán MA, Cole SR. Causal diagrams and measurement bias. Am J Epidemiol 2009;170:959-62.

Herweijer E, Leval A, Ploner A, et al. Association of varying number of doses of quadrivalent human papillomavirus vaccine with incidence of condyloma. JAMA 2014;311:597-603.

Herweijer E, Feldman AL, Ploner A, et al. The participation of HPV-vaccinated women in a national cervical screening program: population-based cohort study. PLoS One 2015;10:e0134185.

Hofmeyr GJ, Abdel-Aleem H, Abdel-Aleem MA. Uterine massage for preventing postpartum haemorrhage. Cochrane Database Syst Rev 2008;3:CD006431.

Holman DM, Benard V, Roland KB, et al. Barriers to human papillomavirus vaccination among US adolescents: a systematic review of the literature. JAMA Pediatr. 2014 Jan;168(1):76-82.

Insinga RP, Liaw K, Johnson LG, et al. A systematic review of the prevalence and attribution of human papillomavirus types among cervical, vaginal, and vulvar precancers and cancers in the United States. Cancer Epidemiol Biomarkers Prev 2008;17:1611-22.

InterAgency Working Group on Costing (IAWG-Costing), Avenir Health. OneHealth Tool. 2015. Available at < http://www.avenirhealth.org/software-onehealth>.

International Institute for Population Sciences (IIPS). District level household and facility aurvey (DLHS-4), 2012-13. 2014. Karnataka, India: IIPS. Available at <https://nrhm-mis.nic.in/DLHS4/State%20Reports/Karnataka.pdf>.

Jain N, Euler GL, Shefer A, et al. Human papillomavirus (HPV) awareness and vaccination initiation among women in the United States, National Immunization Survey-Adult 2007. Prev Med 2009;48:426-31.

Kahn JA, Rosenthal SL, Jin Y, et al. Rates of human papillomavirus vaccination, attitudes about vaccination, and human papillomavirus prevalence in young women. Obstet Gynecol 2008;111:1103-10.

Kahn JA, Brown DR, Ding L, et al. Vaccine-type human papillomavirus and evidence of herd protection after vaccine introduction. Pediatrics 2012;130:e249-56.

Kaiser Family Foundation. Summary of new health reform law – 2011. Available at <http://www.kff.org/healthreform/upload/8061.pdf>.

Kassebaum NJ, Bertozzi-Villa A, Coggeshall MS, et al. Global, regional, and national levels and causes of maternal mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014;384:980-1004.

Kim JJ, Goldie SJ. Cost-effectiveness of including boys in a human papillomavirus vaccination programme in the United States. BJM 2009;339:b3884.

Kjaer SK, Sigurdsson K, Iversen O, et al. A pooled analysis of continued prophylactic efficacy of quadrivalent human papillomavirus (types 6/11/16/18) vaccine against high-grade cervical and external genital lesions. Cancer Prev Res 2009;2:868-78.

Koerkamp BG, Stijnen T, Weinstein MC, et al. The combined analysis of uncertainty and patient heterogeneity in medical decision models. Med Decis Making 2011;31:650-661.

Kreimer AR, Rodriguez AC, Hildesheim A, et al. Proof-of-principle evaluation of the efficacy of fewer than three doses of a bivalent HPV16/18 vaccine. J Natl Cancer Inst 2011;103:1444-51.

Kreimer AR, Struyf F, Del Rosario-Raymundo MR, et al. Efficacy of fewer than three doses of an HPV-16/18 AS04-adjuvanted vaccine: combined analysis of data from the Costa Rica Vaccine and PATRICIA trials. Lancet Oncol 2015;16:775-86.

Kumar V, Abbas AK, Aster JC. “Female genital system and breast,” in Robbins Basic Pathology, 9th edition. 2013. Philadelphia, PA: Saunders.

Lacey CJN, Lowndes CM, Shah KV. Chapter 4: burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease. Vaccine 2006;24(Suppl 3):S3/35-41.

Lehtinen M, Paavonen J, Wheeler CM, et al. Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 2012;13:89-9.

Leyden WA, Manos MM, Geiger AM, et al. Cervical cancer in women with comprehensive health care access: attributable factors in the screening process. J Natl Cancer Inst 2005;97:675-83.

Levin A, Dmytraczenko T, McEuen M, et al. Costs of maternal health care services in three Anglophone African countries. Int J Health Plann Manage 2003;18:3-22.

Liddon NC, Leichliter JS, Markowitz LE. Human papillomavirus vaccine and sexual behavior among adolescent and young women. Am J Prev Med 2012;42:44-52.

Lindley MC, Shen AK, Orenstein WA, et al. Financing the delivery of vaccines to children and adolescents: challenges to the current system. Pediatrics 2009;124 Suppl 5:S548-57.

Marchand E, Glenn BA, Bastani R. HPV vaccination and sexual behavior in a community college sample. J Community Health 2013;38:1010-4.

Markowitz LE, Dunne EF, Saraiya M, et al. Quadrivalent human papillomavirus vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2007;56:1-24.

Markowitz LE, Hariri S, Lin C, et al. Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003-2010. J Infect Dis 2013;208:385-93.

Markowitz LE, Dunne EF, Saraiya M, et al. Human papillomavirus vaccination: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2014;63:1-30.

Maynard A. The powers and pitfalls of payment for performance. Health Econ 2012;21:3-12.

Meacock R, Kristensen SR, Sutton M. The cost-effectiveness of using financial incentives to improve provider quality: a framework and application. Health Econ 2014;23:1-13.

Meessen B, Soucat A, Sekabaraga C. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? Bull World Health Organ 2011;89:153-6.

Miller G, Babiarz KS. Pay-for-Performance incentives in low- and middle-income country health programs. National Bureau of Economic Research Working Paper Series. 2013. Paper 18932.

Mohanan M, Miller G, La Forgia G, et al. Improving maternal and child health in India: evaluating demand and supply side strategies (IMATCHINE), 3ie Grantee Final Report. 2015. New Delhi: International Initiative for Impact Evaluation (3ie). Available at <http://www.3ieimpact.org/media/filer_public/2015/12/03/gfr-ow2205-imatchine.pdf>

Mohanan M, Miller G, Donato K, et al. Input-based versus output-based incentive contracts in health care: experimental evidence form India. Paper forthcoming.

Muñoz N, Kjaer SK, Sigurdsson K, et al. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. J Natl Cancer Inst 2010;102:325-39.

National Conference of State Legislatures. HPV vaccine: State legislation and statutes. Updated 2014. Available at: <http://www.ncsl.org/issues-research/health/hpv-vaccine-state-legislation-and-statutes.aspx>.

Natunen K, Lehtinen J, Namujju P, et al. Aspects of prophylactic vaccination against cervical cancer and other human papillomavirus-related cancers in developing countries. Infect Dis Obstet Gynecol 2011;2011:675858.

Newhouse J, Insurance Experiment Group. Free for all? Lessons from the RAND Health Insurance Experiment. 1993. Cambridge, MA: Harvard University Press.

Newhouse JP, Phelps CE. "Price and income elasticities for medical care services," in Perlman M, ed., The Economics of Health and Medical Care. 1974. London: Macmillan.

Petersen LA, Woodard LD, Urech T, et al. Does pay-for-performance improve the quality of health care? Ann Intern Med 2006;145:265-72.

President's Cancer Panel. Accelerating HPV vaccine uptake: urgency for action to prevent cancer. President's Cancer Panel Annual Report 2012-2013. Available at <http://deainfo.nci.nih.gov/advisory/pcp/annualReports/HPV/PDF/PCP_Annual_Report_2012-2013.pdf>.

Rahman M, Laz TH, Berenson AB. Geographic variation in human papillomavirus vaccination uptake and young adult women in the United States during 2008-2010. Vaccine 2013;31:5495-9.

Reagan-Steiner S, Yankey D, Jeyarajah J, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13-17 years--United States, 2014. MMWR Morb Mortal Wkly Rep 2015;64:784-92.

Reed M, Price M, Brand R, et al. Deductible-based health insurance: sharpening a blunt instrument. Health Aff 2009;28:1145-54.

Registrar General of India, Centre for Global Health Research. Maternal mortality in India, 1997-2003: Trends, causes and risk factors. 2006. Available at <http://www.cghr.org/wordpress/wp-content/uploads/RGI-CGHR-Maternal-Mortality-in-India-1997–2003.pdf>.

Riffe TLM. The maternal mortality incidence rate. 2010. Paper presented at the 2011 Annual Meeting of the Population Association of America. Available at <http://paa2011.princeton.edu/papers/110626>.

Ringel JS, Hosek SD, Vollaard BA, et al. The Elasticity of Demand for Health Care: A Review of the Literature and Its Application to the Military Health System. 2002. Santa Monica, CA: RAND.

Rosett RN, Huang LF. The effect of health insurance on the demand for medical care. Journal of Political Economy 1973;81:281-305.

Rowe JW, Brown-Stevenson T, Downey RL, et al. The effect of consumer-directed health plans on the use of preventive and chronic illness services. Health Aff 2008;27:113-20.

Safaeian M, Porras C, Pan Y, et al. Durable antibody responses following one dose of the bivalent human papillomavirus L1 virus-like particle vaccine in the Costa Rica Vaccine Trial. Cancer Prev Res (Phila) 2013;6:1242-50.

Saslow D, Runowicz CD, Solomon D, et al. American Cancer Society guideline for the early detection of cervical neoplasia and cancer. CA Cancer J Clin 2002;52:342-62.

Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin 2012;62:147-72.

Schiller JT, Castellsague X, Garland SM. A review of clinical trials of human papillomavirus prophylactic vaccines. Vaccine 2012;30(Suppl 5):F123-38.

Scitovsky AA, McCall NM. Coinsurance and the demand for physician services. Social Security Bulletin 1977;40:19-27.

Shahar E. Causal diagrams for encoding and evaluation of information bias. J Eval Clin Pract 2009;15:436-440.

Smith PJ, Lindley MC, Shefer A, et al. Underinsurance and adolescent immunization delivery in the United States. Pediatrics 2009;124(suppl 5):S515-21.

Solomon D, Davey D, Kurman R, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA 2002;287:2114-9.

Souza JP, Cecatti JG, Pacagnella RC, et al. Development and validation of a questionnaire to identify severe maternal morbidity in epidemiological surveys. Reprod Health 2010;7:16.

Stewart MK, Festin M. Validation study of women's reporting and recall of major obstetric complications treated at the Philippine General Hospital. Int J Gynaecol Obstet 1995;48 Suppl:S53-66.

Stokley S, Jeyarajah J, Yankey D, et al. Human papillomavirus vaccination coverage among adolescents, 2007–2013, and postlicensure vaccine safety monitoring, 2006–2014—United States. MMWR 2014;63:620-4.

Tielsch JM. The know, do, and quality gaps in international maternal and child health care interventions. JAMA Pediatr 2015;169:313-4.

U.S. Preventive Services Task Force. Screening for cervical cancer: recommendations and rationale. 2003. Rockville, MD: Agency for Healthcare Research and Quality.

U.S. Preventive Services Task Force. Screening for cervical cancer. 2012. Available at <http://www.uspreventiveservicestaskforce.org/uspstf11/cervcancer/cervcancerrs.htm>.

Walboomers JMM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12-9.

Weinstock H, Berman S, Cates W Jr. Sexually transmitted diseases among American youth: Incidence and prevalence estimates, 2000. Perspect Sex Rerod Health 2004;36:6-10.

Weissman E, Sentumbe-Mugisa O, Mbonye AK, et al. Uganda safe motherhood programme costing study. 1999. Available at <http://www.who.int/reproductivehealth/publications/maternal_perinatal_health/RHR_99_9/en/>.

Wheeler CM, Castellsague X, Garland SM, et al. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 2012;13:100-10.

Williams WW, Lu PJ, Saraiya M, et al. Factors associated with human papillomavirus vaccination among young adult women in the United States. Vaccine 2013;31:2937-46.

Witter S, Fretheim A, Kessy FL, et al. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. Cochrane Database Syst Rev 2012;2:CD007899.

Workowski KA, Bolan GA. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recomm Rep 2015;64:1-137.

World Bank. Gross domestic product per capita by country. 2015. Available at <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.

World Health Organization, United Nations Population Fund, United Nations Children's Fund. Guidelines for monitoring the availability and use of obstetric services. 1997. New York: UNICEF, WHO, UNFPA.

World Health Organization. Managing complications in pregnancy and childbirth: a guide for midwives and doctors. 2003. Geneva: WHO.

World Health Organization. Managing puerperal sepsis. 2008. Available at <http://apps.who.int/iris/bitstream/10665/44145/6/9789241546669_6_eng.pdf>.

World Health Organization. WHO Recommended Interventions for Improving Maternal and Newborn Health. 2009. Available at <http://whqlibdoc.who.int/hq/2007/WHO_MPS_07.05_eng.pdf>.

World Health Organization. WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia. 2011. Available at <http://apps.who.int/iris/bitstream/10665/44703/1/9789241548335_eng.pdf>.

World Health Organization. WHO recommendations for the prevention and treatment of postpartum haemorrhage. 2012. Available at <http://apps.who.int/iris/bitstream/10665/75411/1/9789241548502_eng.pdf>.

World Health Organization. Life tables by country: India. 2013. Available at <http://apps.who.int/gho/data/view.main.60740?lang=en>.

World Health Organization. Human papillomavirus vaccines: WHO position paper, October 2014. Wkly Epidemiol Rec 2014;89:465-91.

World Health Organization. Trends in maternal mortality: 1990 to 2015: estimates by WHO, UNICEF, UNFPA, the World Bank and the UN Population Division. 2015a. Available at <http://apps.who.int/iris/bitstream/10665/194254/1/9789241565141_eng.pdf?ua=1>.

World Health Organization. Threshold values for intervention cost-effectiveness by region. 2015b. Available at < http://www.who.int/choice/costs/CER_levels/en/ >.

World Health Organization, UNICEF, United Nations Fund for Population Activities, World Bank Group, United Nations Population Division. Maternal mortality in 1990-2015: India. 2015c. Available at <http://www.who.int/gho/maternal_health/countries/ind.pdf>.

World Health Organization. Choosing interventions that are cost-effective (WHO-CHOICE): Quantities and unit prices. 2015d. Available at <http://www.who.int/choice/cost-effectiveness/inputs/en/>.

Wright TC Jr, Massad LS, Dunton CJ, et al. 2006 consensus guidelines for the management of women with abnormal cervical screening tests. J Low Genit Tract Dis 2007;11:201-22.

APPENDIX A: Supplemental material for Paper 1

A.1: Procedures codes used for sample selection or variable measurement

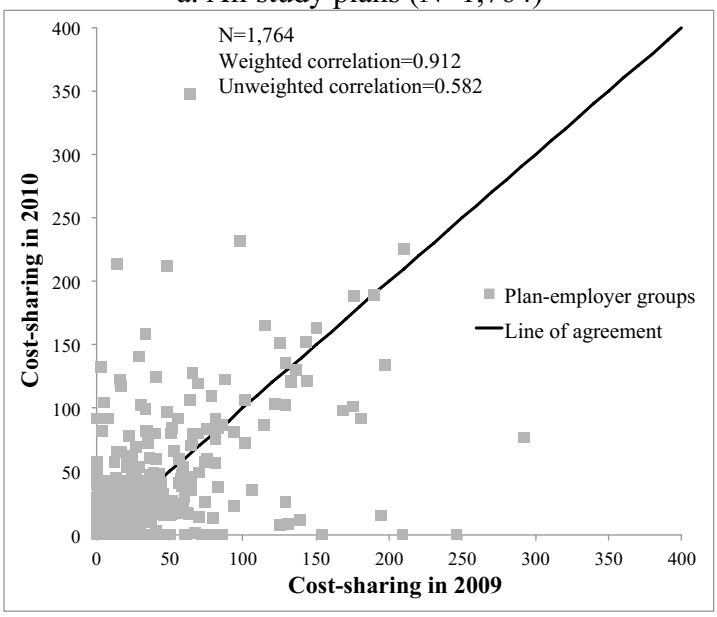
Table A.1: List of procedure codes

Code type / Code	Category / Description
CPT or NDC	***Procedure or drug codes for HPV vaccines***
90649	Human Papilloma virus (HPV) vaccine, types 6, 11, 16, 18 (quadrivalent), 3 dose schedule, for intramuscular use
90650	Human Papilloma virus (HPV) vaccine, types 16, 18, bivalent, 3 dose schedule, for intramuscular use
00006-4045-xx	Human Papillomavirus Quadrivalent (Types 6, 11, 16, and 18) Vaccine, Recombinant
58160-0830-xx	Human Papillomavirus Bivalent (Types 16 and 18) Vaccine, Recombinant
CPT, HCPCS or ICD-9	***Procedure codes for wellness visits and vaccine administration***
99201-99205, 99211-99215, 99241-99245	Office or other outpatient visit for the evaluation and management of a new or established patient
99381-99385, 99391-99395	Initial or periodic comprehensive preventive medicine evaluation and management of an individual
99401-99404, 99411-99412	Preventive medicine counseling and/or risk factor reduction intervention(s) provided to an individual or individuals in a group setting
99420	Administration and interpretation of health risk assessment instrument (eg, health hazard appraisal)
99429	Unlisted preventive medicine service
S0610, S0612, S0613	Annual gynecological examination
90471, 90472, 90772, 96372	Vaccine administration
V04.89	Need for prophylactic vaccination and inoculation against other viral diseases
V05.8, V05.9	Need for prophylactic vaccination and inoculation against other specified or unspecified disease
V06.8, V06.9	Need for prophylactic vaccination and inoculation against other combinations of diseases
V20.2	Routine infant or child health check
V70.0	Routine general medical examination at a health care facility
V72.31	Routine gynecological examination
V76.2	Screening for malignant neoplasms of cervix

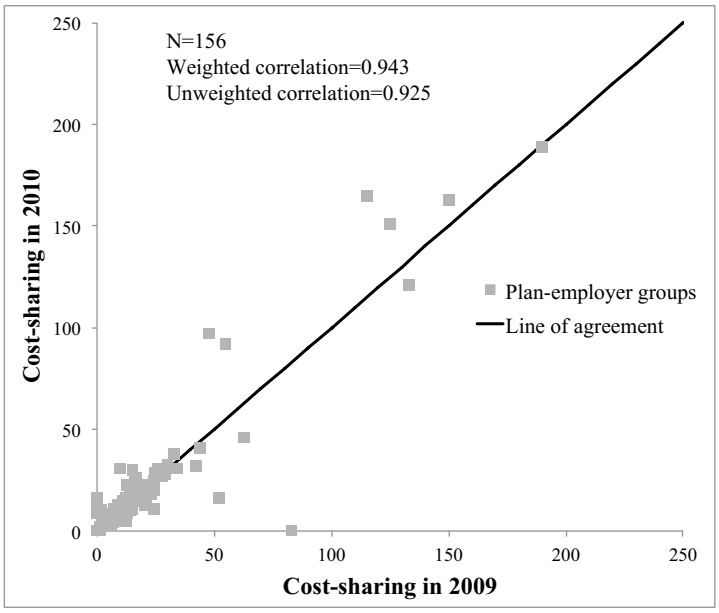
A.2: Concordance analysis to test the precision of empirical cost-sharing estimates

Figure A.1: Scatterplot: Estimated cost-sharing per HPV dosage in 2009 vs. 2010

a. All study plans (N=1,764)



b. Large plans only (N=156)



The weighted Pearson correlation coefficients are weighted by plans' sample size in 2009. Cost-sharing in either year is inflation-adjusted to 2013 USD. In Figure A.1a, the weighted correlation of 0.912 indicates high concordance between 2009 and 2010 cost-sharing estimates in the overall sample. The lower unweighted correlation of 0.582 indicates worse concordance among smaller plan-employer groups. Figure A.1b focuses on large plans only, proxied by having ≥1,000 study subjects enrolled in 2009; this subgroup comprises 77% of the individual-level sample (in terms of person-years of enrollment). The scatterplots confirm that the concordance of estimates is especially high in large plans, suggesting that cost-sharing levels are precisely estimated in these plans.

A.3: Validations of empirical cost-sharing estimates against MarketScan Summary Plan Description Extractions

Additional validations are performed using the subset of study plans that appear in the Extractions database for any policy year(s) during 2010-2013. The purpose of these analyses is to verify that empirical cost-sharing estimates systematically differ between plans with versus without cost-sharing for routine vaccinations (as stated in plan materials), and between plans that did versus did not claim grandfathered status.

Among plans that stated zero cost-sharing for annual wellness visits and routine vaccinations in a given year, plan-level cost-sharing estimates are $4.75 per dosage on average, and are ≤$10 per dosage for the vast majority (96%) of plans (Table A.2). Among plans that indicated non-zero cost-sharing for either/both of these services, cost-sharing estimates are $27.75 on average and ≤$10 for 44% of plans. Most non-zero cost-sharing plans with estimates ≤$10 had merely stated that cost-sharing *may* apply for the office visit if billed separately from the vaccination. Cost-sharing estimates are generally higher in plans that formally imposed cost-sharing for wellness visits (mean: $15.73; percent ≤$10: 22%), or had incomplete coverage for both vaccination itself and wellness visits (mean: $94.40; percent ≤$10: 25%).

Table A.2: Empirical cost-sharing statistics by level of coverage stated in plan materials (2010-2013)

Stated level of coverage for routine vaccinations and annual wellness visits	Plan-years	Person-years of enrollment	Empirical cost-sharing (USD 2013)		
			Average	(SD)	Percent ≤$10
Free	402	310,960	4.75	(3)	96%
Ambiguous	75	37,026	8.39	(6)	66%
Not free	133	103,072	27.75	(44)	44%
Not free - Office visit cost-sharing may apply	*50*	*37,376*	*9.32*	*(5)*	*81%*
Not free - Office visit cost-sharing applies	*50*	*46,901*	*15.73*	*(5)*	*22%*
Not free - Coverage incomplete for other reasons	*33*	*18,795*	*94.40*	*(69)*	*25%*

Table includes study plans appearing in the Extractions database for one or more calendar years during 2010-2013. The "plan-years" column counts the number of unique plan and calendar year combinations categorized into each level of coverage. To compute the empirical cost-sharing statistics, the cost-sharing estimate for each plan/year is weighted by the number of study subjects enrolled in that calendar year. Each plan/year is categorized into the following coverage levels based on statements in plan materials:
"Free": Routine vaccinations and annual wellness visits are fully covered.
"Ambiguous": Annual wellness visits or preventive services are fully covered, but vaccine coverage is not specified.
"Not free - office visit cost-sharing may apply": Routine vaccinations are fully covered, but cost-sharing may be imposed for the office visit if the visit is billed separately.
"Not free - office visit cost-sharing applies": Routine vaccinations are fully covered after an office visit charge.
"Not free - coverage incomplete for other reasons": Other coverage limitations applying to both vaccinations and the office visit (e.g., coinsurance applying to both, deductible without exemptions, annual preventive care limit, etc.).

Grandfathered status (yes/no) could be determined for 251 plan-years during 2011-2013, based on the presence/absence of a grandfathered status disclosure in the full plan benefit booklet (Table A.3). Among plans that claimed to be grandfathered (25 plan-years), empirical cost-sharing is $67.59 on average (percent ≤$10: 26%). A minority of grandfathered plans (13%) still claimed to fully cover routine vaccinations and annual wellness visits. However, empirical cost-sharing levels are somewhat higher in these plans (mean: $7.96; percent ≤$10: 82%) than in plans without a grandfathered status disclosure (mean: $4.76; percent ≤$10: 97%).

Table A.3: Empirical cost-sharing statistics by grandfathered status & level of coverage stated in plan materials (2011-2013; Full booklet extractions only)

Stated grandfathered status / Stated level of coverage for routine vaccinations and annual wellness visits	Plan-years	Person-years of enrollment	Empirical cost-sharing (USD 2013)		
			Average	(SD)	Percent ≤$10
Non-grandfathered plans	226	228,599	4.76	(3)	97%
Free	*194*	*215,217*	*4.66*	*(2)*	*97%*
Ambiguous	*26*	*8,847*	*7.16*	*(6)*	*93%*
Not free	*6*	*4,535*	*5.21*	*(2)*	*98%*
Grandfathered plans	25	31,055	62.59	(67)	26%
Free	*10*	*4,140*	*7.96*	*(2)*	*82%*
Ambiguous	*1*	*28*	*20.00*	*(0)*	*0%*
Not free	*14*	*26,887*	*71.05*	*(68)*	*18%*

Table includes study plans appearing in the Extractions database for one or more calendar years during 2011-2013, excluding plan-years in which the extracted data was based on a benefit table or summary rather than the full plan booklet. (The full plan booklet is needed to reliably determine the presence/absence of a grandfathered status claim.) See previous table footnotes for additional details on the empirical cost-sharing statistics and stated levels of coverage.

A.4: Testing alternative functional forms of the cost-sharing policy variable

Below, Table A.4 summarizes results from models of time to vaccine initiation using non-linear functions of cost-sharing level, including: a log-transformation of cost-sharing plus $1; a quadratic function of cost-sharing; and a dichotomous indicator of "free/near-free" versus "not free" vaccination:

Table A.4: Hazard ratios of vaccine initiation and changes in model fit when using alternative functional forms of the cost-sharing variable

Functional form of policy variable	Hazard ratio per unit reduction (95% CI)	Change in -2 log L (vs. row [i])
Linear cost-sharing variable (main analysis):		
(i) $cs_{ip}(t)$	1.041*** (1.033, 1.048)	-
Log-transformation of cost-sharing plus $1:		
(ii) $\ln[cs_{ip}(t)+0.1]$	1.024*** (1.011, 1.037)	+98.7
Quadratic function of cost-sharing:		
(iii) $cs_{ip}(t)$	1.033*** (1.019, 1.047)	-1.7
$cs_{ip}(t)^2$	1.0007 (1.000, 1.002)	
Dichotomous indicator of "free/near-free" or "not free" (using $10 cutoff):		
(iv) $I[cs_{ip}(t) > 1]$	1.026** (1.010, 1.041)	+102.5

Note that the linear cost-sharing variable $cs_{ip}(t)$ is equal to the cost-sharing level (in USD 2013) divided by 10 that individual *i* faces in plan *p* at age 9+*t* years. For each cost-sharing variable in the leftmost column, the hazard ratio is for a 1-unit reduction in the transformed cost-sharing variable; for example, when using the dichotomous cost-sharing variable, the hazard ratio of vaccine initiation is 1.026 for cost-sharing levels ≤$10 (vs. levels >$10).
**p<0.05; **p<0.01; ***p<0.001*

Model fit does not significantly improve when a squared cost-sharing term was added to the model (based on the decrease of 1.7 in -2 log L on 1 d.f.; chi-square p=0.192), and worsens when the linear variable was replaced with a log-transformation (increase of 98.7 in -2 log L) or with a dichotomous variable (increase of 102.5 in -2 log L). These results support the use of a

linear price variable when modeling demand for HPV vaccination via a Cox proportional hazard model.

Note that, when using a dichotomous policy variable, the effect size is considerably muted compared to the main effect estimate based on a continuous policy variable. Implicitly, using a binary policy variable assumes that any reduction in cost-sharing that does not reach the "free/near-free" threshold (e.g., $10) is unrelated to the preventive care reforms: Plan-employer groups may have large reductions in cost-sharing in post-ACA years, but are still treated as part of the control group if estimated cost-sharing levels remain above the threshold. Consequently, this approach yields a very conservative estimate of the overall VBID effect among study plans. It is likely that even incomplete cost-sharing reductions are directly or indirectly attributable to the preventive care reforms (e.g., due to partial compliance among grandfathered plans, caveats to free preventive care in non-grandfathered plans, or random measurement error in the cost-sharing estimates).

APPENDIX B: Supplemental material for Paper 2

B.1: Diagnosis and procedures codes used for sample selection or variable measurement

Table B.1: List of diagnosis and procedure codes

Code type / Code	Category / Description
CPT or NDC	***Procedure or drug codes for HPV vaccines***
90649	Human Papilloma virus (HPV) vaccine, types 6, 11, 16, 18 (quadrivalent), 3 dose schedule, for intramuscular use
90650	Human Papilloma virus (HPV) vaccine, types 16, 18, bivalent, 3 dose schedule, for intramuscular use
00006-4045-xx	Human Papillomavirus Quadrivalent (Types 6, 11, 16, and 18) Vaccine, Recombinant
58160-0830-xx	Human Papillomavirus Bivalent (Types 16 and 18) Vaccine, Recombinant
ICD-9	***Diagnosis codes for cytological abnormalities***
795.00	Abnormal glandular Papanicolaou smear of cervix (Applies to atypical endocervical, endometrial, or cervical glandular cells NOS)
795.01	Papanicolaou smear of cervix with atypical squamous cells of undetermined significance (ASC-US)
795.02	Papanicolaou smear of cervix with atypical squamous cells cannot exclude high grade squamous intraepithelial lesion (ASC-H)
795.03	Papanicolaou smear of cervix with low grade squamous intraepithelial lesion (LGSIL)
795.04	Papanicolaou smear of cervix with high grade squamous intraepithelial lesion (HGSIL)
795.06	Papanicolaou smear of cervix with cytologic evidence of malignancy
ICD-9	***Diagnosis codes for histology grade***
622.10	Dysplasia of cervix, unspecified (Applies to anaplasia of cervix, cervical atypism, or cervical dysplasia NOS)
622.11	Mild dysplasia of cervix (Applies to cervical intraepithelial neoplasia I [CIN I])
622.12	Moderate dysplasia of cervix (Applies to CIN II)
233.1	Carcinoma in situ of cervix uteri (Applies to CIN III and adenocarcinoma in situ of cervix)
180.x	Malignant neoplasm of cervix uteri
CPT, HCPCS or ICD-9	***Procedure codes for Pap cytology procedures***
88141-88145, 88147-88148, 88150, 88152-88155, 88164-88167, 88174-88175	Cytopathology, cervical or vaginal
G0101	Cervical or vaginal cancer screening; pelvic and clinical breast examination
G0123-G0124	Screening cytopathology, cervical or vaginal (any reporting system), collected in preservative fluid, automated thin layer preparation
G0141	Screening cytopathology smears, cervical or vaginal, performed by automated system, with manual rescreening, requiring interpretation by physician
G0143	Screening cytopathology, cervical or vaginal (any reporting system), collected in preservative fluid, automated thin layer preparation; with manual screening and rescreening by cytotechnologist under physician supervision

Table B.1 (Continued)

G0144	Screening by automated system, under physician supervision
G0145	Screening cytopathology smears, cervical or vaginal, performed by automated system, with manual rescreening, requiring interpretation by physician
G0147	Screening cytopathology smears, cervical or vaginal; performed by automated system under physician supervision
G0148	Performed by automated system with manual rescreening
Q0091	Screening Papanicolaou smear; obtaining, preparing and conveyance of cervical or vaginal smear to laboratory
V72.32	Encounter for Papanicolaou cervical smear to confirm findings of recent normal smear following initial abnormal smear
V76.2	Screening for malignant neoplasms of cervix
CPT or HCPCS	***Procedure codes for histology follow-up***
57420	Colposcopy of the entire vagina, with cervix if present
57452	Colposcopy of the cervix including upper/adjacent vagina
57421	Colposcopy of the entire vagina, with cervix if present; with biopsy(s) of vagina/cervix
57455	Colposcopy of the cervix including upper/adjacent vagina; with biopsy(s) of the cervix
57500	Biopsy, single or multiple, or local excision of lesion, with or without fulguration (separate procedure)
57505	Endocervical curettage
57454	Colposcopy of the cervix including upper/adjacent vagina; with biopsy(s) of the cervix and endocervical curettage
57456	Colposcopy of the cervix including upper/adjacent vagina; with endocervical curettage
57450	Colposcopy of the cervix including upper/adjacent vagina; with loop electrode biopsy(s) of the cervix
57460	Colposcopy of the cervix including upper/adjacent vagina; with loop electrode biopsy(s) of the cervix
57461	Colposcopy of the cervix including upper/adjacent vagina; with loop electrode conization of the cervix
57520	Conization of cervix, including cold knife or laser
57522	Conization of cervix, including cold knife or laser
57510, 57511, 57513	Other surgical cautery of cervix
CPT or ICD-9	***Procedure and diagnosis codes related to HPV DNA tests***
87620	Infectious agent detection by nucleic acid (DNA or RNA); papillomavirus, human, direct probe technique
87621	Infectious agent detection by nucleic acid (DNA or RNA); papillomavirus, human, amplified probe technique
87622	Infectious agent detection by nucleic acid (DNA or RNA); papillomavirus, human, quantification
V73.81	Special screening examination for Human papillomavirus (HPV)
795.05	High-risk HPV DNA test positive
795.09	Other abnormal Papanicolaou smear of cervix and cervical HPV (Applies to cervical low risk HPV DNA test positive)
CPT or ICD-9	***Procedure codes for chlamydia or gonorrhea testing***
86631	Antibody; Chlamydia

Table B.1 (Continued)

86632	Antibody; Chlamydia, IgM
87110	Culture, chlamydia, any source
87270	Infectious agent antigen detection by immunofluorescent technique; Chlamydia trachomatis
87320	Infectious agent antigen detection by enzyme immunoassay technique, qualitative or semiquantitative, multiple-step method; Chlamydia trachomatis
87490	Infectious agent detection by nucleic acid (DNA or RNA); Chlamydia trachomatis, direct probe technique
87491	Infectious agent detection by nucleic acid (DNA or RNA); Chlamydia trachomatis, amplified probe technique
87492	Infectious agent detection by nucleic acid (DNA or RNA); Chlamydia trachomatis, quantification
87810	Infectious agent antigen detection by immunoassay with direct optical observation; Chlamydia trachomatis
86729	Antibody; lymphogranuloma venereum
87850	Infectious agent antigen detection by immunoassay with direct optical observation; Neisseria gonorrhoeae
87590	Infectious agent detection by nucleic acid (DNA or RNA); Neisseria gonorrhoeae, direct probe technique
87591	Infectious agent detection by nucleic acid (DNA or RNA); Neisseria gonorrhoeae, amplified probe technique
87592	Infectious agent detection by nucleic acid (DNA or RNA); Neisseria gonorrhoeae, quantification
V73.88	Special screening examination for other specified chlamydial diseases
V73.98	Special screening examination for unspecified chlamydial disease
V74.5	Screening examination for venereal disease
ICD-9	***Diagnosis codes for chlamydia or gonorrhea***
091.xx-097.xx	Syphilis diseases
098.xx	Gonococcal infections

Notes: I search both outpatient and inpatient medical visit records for claims associated with relevant procedure and diagnosis codes. I also search prescription drug records for any HPV vaccine doses delivered through outpatient pharmacies, although the vast majority of vaccinations in the cohort occurred during medical visits.
CPT, Current Procedural Terminology; HCPCS, Healthcare Common Procedure Coding System; ICD-9, International Classification of Diseases, Ninth Revision; NDC, National Drug Code.

B.2: Suitability of claims-based risk proxies

The main analyses use history of testing and diagnosis for chlamydia or gonorrhea as claims-based proxies for background HPV exposure risk. These conditions are sexually-transmitted diseases for which national guidelines recommend routine screening in sexually-active girls and young women (Workowski 2015). To assess the suitability of either proxy, I fit a logistic regression model to estimate their independent associations with risk of any cytological abnormality at the first Pap screening, controlling for vaccine dose level, all demographic and socioeconomic covariates, and indicators for all combinations of age and calendar year at the first screening. Regression output is presented in Table B.2; as shown, both variables are strongly and independently associated with increased cytological abnormality risk (both $p<0.001$).

Table B.2: Associations between study covariates and risk of any cytological abnormality at first Pap screening: Multivariate logistic regression results

Covariate	Odds ratio	(95% CI)	
Claim(s) at or before first Pap (yes vs. no)			
Chlamydia or gonorrhea testing	1.362	(1.325, 1.400)	***
Chlamydia or gonorrhea diagnosis	1.858	(1.631, 2.117)	***
Census division (vs. South Atlantic)			
East North Central	0.970	(0.929, 1.012)	
East South Central	1.132	(1.072, 1.197)	***
Middle Atlantic	0.911	(0.856, 0.970)	**
Mountain	0.924	(0.860, 0.992)	*
New England	0.990	(0.910, 1.077)	
Pacific	0.738	(0.701, 0.778)	***
West North Central	1.038	(0.974, 1.106)	
West South Central	0.987	(0.947, 1.030)	
Outpatient service use in 2007 (yes vs. no)	0.966	(0.930, 1.004)	
Data contributor type (insurer vs. employer)	1.118	(1.084, 1.154)	***
Plan type (vs. preferred provider organization)			
Comprehensive	1.156	(1.062, 1.259)	***
Exclusive provider organization	0.796	(0.671, 0.944)	**
Health maintenance organization	0.929	(0.894, 0.965)	***
Point-of-service (non-capitated)	1.030	(0.987, 1.075)	
Point-of-service (with capitation)	1.194	(1.030, 1.385)	*
Consumer-driven health plan	0.885	(0.804, 0.975)	*
Unknown/other	0.895	(0.807, 0.993)	*
Socioeconomic measures by zip code			
Median 4-person family income (vs. >400% FPL)			
≤200% FPL	1.157	(1.085, 1.233)	***
201-300% FPL	1.142	(1.089, 1.199)	***
301-400% FPL	1.081	(1.038, 1.126)	***
Quartiles, % white (vs. 4th quartile)			
1 (lowest % white)	1.202	(1.150, 1.257)	***
2	1.071	(1.029, 1.116)	***
3	1.060	(1.021, 1.102)	**
Quartiles, % without high school education (vs. 4th quartile)			
1 (highest educational attainment)	0.887	(0.838, 0.938)	***
2	0.955	(0.909, 1.002)	
3	0.962	(0.922, 1.003)	

Notes: In addition to the covariates shown, the model controls for vaccination status (zero, one, two, or three doses) and for each unique combination of age and calendar year at the first Pap screening.
*p<0.05; **p<0.01; ***p<0.001

B.3: Survival analysis of time to screening initiation

Table B.3: Cox proportional hazards model of time to first Pap cytology screening

Covariate	Hazard ratio	(95% CI)	
Number of HPV vaccine doses (vs. 3 doses)			
0 dose	0.75	(0.75, 0.76)	***
1 dose	0.99	(0.98, 1.01)	
2 dose	0.98	(0.97, 1.00)	
Census division (vs. South Atlantic)			
East North Central	0.95	(0.94, 0.96)	***
East South Central	1.08	(1.06, 1.10)	***
Middle Atlantic	0.87	(0.86, 0.89)	***
Mountain	0.94	(0.92, 0.96)	***
New England	0.77	(0.75, 0.79)	***
Pacific	0.73	(0.72, 0.74)	***
West North Central	0.99	(0.97, 1.01)	
West South Central	1.10	(1.09, 1.12)	***
Outpatient service use in 2007 (yes vs. no)	1.57	(1.55, 1.59)	***
Data contributor type (insurer vs. employer)	0.95	(0.94, 0.96)	***
Plan type (vs. preferred provider organization)			
Comprehensive	0.93	(0.91, 0.96)	***
Exclusive provider organization	1.02	(0.97, 1.07)	
Health maintenance organization	1.02	(1.01, 1.03)	***
Point-of-service (non-capitated)	1.06	(1.05, 1.07)	***
Point-of-service (with capitation)	1.04	(1.00, 1.10)	
Consumer-driven health plan	0.95	(0.92, 0.97)	***
Unknown/other	0.95	(0.92, 0.98)	***
Socioeconomic measures by zip code			
Median family income (4-person) (vs. >400% FPL)			
≤200% FPL	1.14	(1.12, 1.16)	***
201-300% FPL	1.13	(1.12, 1.15)	***
301-400% FPL	1.09	(1.08, 1.10)	***
Quartiles, % white (vs. 4th quartile)			
1 (lowest % white)	0.90	(0.89, 0.91)	***
2	0.94	(0.93, 0.95)	***
3	0.97	(0.95, 0.98)	***
Quartiles, % without high school education (vs. missing socioeconomic measures)			
1 (highest educational attainment)	0.95	(0.92, 0.97)	***
2	1.04	(1.01, 1.07)	*
3	1.05	(1.02, 1.08)	**
4	1.02	(0.99, 1.06)	

Notes: This regression analysis is conducted among 1,226,763 adolescents 9-17 years old in 2007 who were continuously enrolled in MarketScan through 2008 or later and screening-naïve as of January 1, 2008. In addition to the covariates shown, the regression controls for starting age in 2007 by stratifying the baseline hazard function by birth year. Vaccination status corresponds to the number of doses received up to and including each month of follow-up.
*p<0.05; **p<0.01; ***p<0.001

B.4: Summary of vaccine efficacy data from trials

Table B.4: Reported efficacy of quadrivalent and bivalent HPV vaccines against CIN irrespective of HPV type

Vaccine type, Trial name, Months of follow-up, First author (year)	Endpoint	Population[a]	N[b]	Efficacy (95% CI)[c]
Quadrivalent vaccine				
FUTURE I/II, Month 42, Muñoz (2010)	CIN1, all-cause	ITT, all ages	17,160	20.3% (12.4, 27.5)
		HPV-naïve, all ages	9,296	29.7% (16.9, 40.6)
	CIN2+, all-cause	ITT, all ages	17,160	19.0% (7.7, 28.9)
		HPV-naïve, all ages	9,296	42.7% (23.7, 57.3)
	CIN3+, all-cause	ITT, all ages	17,160	18% (2, 31)
		HPV-naïve, all ages	9,296	43.0% (13.0, 63.2)
Bivalent vaccine				
PATRICIA, Month 48, Lehtinen (2012)	CIN2+, all-cause	ITT, all ages	17,402	33.1% (22.2, 42.6)
		HPV-naïve, all ages	10,918	64.9% (52.7, 74.2)
	CIN3+, all-cause	ITT, all ages	17,402	45.6% (28.8, 58.7)
		HPV-naïve, all ages	10,918	93.2% (78.9, 98.7)

[a] Intent-to-treat (ITT) populations in the above trials included a mixed of HPV-exposed and -unexposed women (15-26 years-old at vaccine initiation; mean age: 20 years) who received ≥1 dose. Completion of the three-dose vaccine regimen was high in the ITT populations of both FUTURE I/II (97%) and PATRICIA (92%). The generally HPV-naïve populations included the ITT subset that tested negative to 14 HPV subtypes at day 1.
[b] N refers to the combined number of individuals in the vaccine and control groups.
[c] Vaccine efficacy was defined as (1 - relative risk)*100%, where relative risk refers to the ratio of incidence rates (events per person-years at risk) in the vaccine versus placebo groups. Efficacy against CIN1 was not reported in Lehtinen (2012).

APPENDIX C: Supplemental material for Paper 3

C.1: Logistic regression analysis of maternal complication risks

Table C.1: Additional output from logistic regressions of maternal complication risk

Variable	Odds ratio (95% CI), by dependent variable		
	Postpartum hemorrhage	Sepsis	Pre-Eclampsia
Provider randomization group			
Input P4P	0.61 (0.30, 0.99)	1.56 (0.61, 3.30)	1.15 (0.52, 2.55)
Outcome P4P	0.95 (0.59, 1.59)	1.65 (0.87, 3.26)	1.30 (0.69, 2.89)
Control	(ref)	(ref)	(ref)
Demographics			
Mother's age (years)	0.96 (0.80, 1.21)	0.93 (0.69, 1.49)	0.97 (0.77, 1.27)
Mother's age (years), squared	1.0003 (0.9961, 1.0038)	1.0006 (0.9911, 1.0061)	1.0006 (0.9955, 1.0048)
Mother's education level			
Illiterate	1.46 (0.80, 2.37)	0.79 (0.51, 1.73)	0.41 (0.20, 0.84)
Secondary or higher	1.08 (0.78, 1.60)	0.79 (0.48, 1.43)	1.03 (0.70, 1.62)
Primary only	(ref)	(ref)	(ref)
Household's caste			
General or other	1.04 (0.68, 1.58)	1.01 (0.52, 2.01)	1.17 (0.73, 1.74)
Scheduled caste	1.05 (0.63, 1.75)	1.49 (0.72, 3.18)	1.09 (0.69, 1.73)
Other backward class	1.13 (0.74, 1.93)	1.20 (0.50, 2.70)	1.18 (0.69, 1.85)
Scheduled tribe	(ref)	(ref)	(ref)
Household owns land, %	0.87 (0.70, 1.05)	0.89 (0.62, 1.27)	0.97 (0.73, 1.27)
Household has Below Poverty Line card, %	1.16 (0.92, 1.51)	1.11 (0.80, 1.60)	0.97 (0.73, 1.19)
Clinical characteristics			
Mother's first pregnancy	1.18 (0.90, 1.55)	1.54 (1.00, 2.70)	1.06 (0.75, 1.51)
Number of previous children birthed	1.02 (0.87, 1.19)	1.25 (1.00, 1.67)	1.01 (0.86, 1.22)
Mother has had a stillbirth or abortion	1.20 (0.80, 1.68)	1.16 (0.62, 2.03)	0.92 (0.62, 1.43)
Comorbidity history			
Hypertension	1.48 (0.88, 2.46)	0.78 (0.24, 1.30)	1.83 (0.81, 2.76)
Diabetes	1.19 (0.44, 3.38)	1.55 (0.10, 6.78)	0.86 (0.21, 1.83)
Previous stomach surgery	1.11 (0.85, 1.41)	1.05 (0.57, 1.79)	1.00 (0.61, 1.43)
Provider characteristics			
Female provider	0.91 (0.54, 1.39)	0.79 (0.37, 1.50)	1.13 (0.57, 2.11)
Credentials			
MBBS	2.71 (1.19, 8.30)	1.16 (0.33, 13.88)	3.16 (1.43, 12.75)
BAMS	1.48 (0.48, 5.23)	0.70 (0.13, 9.70)	1.06 (0.22, 6.47)
Other qualification	(ref)	(ref)	(ref)
Years practicing	1.01 (0.98, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)
N	2,608	2,608	2,608

Notes: Logistic regression estimates are based on the household survey sample from the experiment. In addition to the covariates listed above, regressions also included indicators for the district in which the provider's facility is located. Confidence intervals are from cluster bootstrapping and are constructed using the percentile method.

C.2: Calculation of age-specific life expectancies

Table C.2: Life table for females in India, 2013

Interval [t, t+1)	Morality rate in [t, t+1) h(t)	Cumulative hazard at t+1 H(t+1)	Survival at t+1 S(t+1)	Life expectancy from t LE(t)	Discounted life expectancy from t disc_LE(t)
[0, 1)	0.044016	0.044016	0.956939	68.126950	29.750059
[1, 2)	0.003563	0.047579	0.953535	70.170100	30.018607
[2, 3)	0.003563	0.051142	0.950144	69.418777	29.921515
[3, 4)	0.003563	0.054705	0.946764	68.664773	29.822121
[4, 5)	0.003563	0.058268	0.943397	67.908077	29.720363
[5, 6)	0.001162	0.059430	0.942301	67.148680	29.615559
[6, 7)	0.001162	0.060592	0.941207	66.226171	29.485041
[7, 8)	0.001162	0.061754	0.940114	65.302589	29.350781
[8, 9)	0.001162	0.062916	0.939022	64.377934	29.212667
[9, 10)	0.001162	0.064078	0.937932	63.452203	29.070583
[10, 11)	0.000859	0.064937	0.937127	62.525396	28.924411
[11, 12)	0.000859	0.065796	0.936322	61.578699	28.770927
[12, 13)	0.000859	0.066655	0.935518	60.631188	28.612964
[13, 14)	0.000859	0.067514	0.934715	59.682863	28.450387
[14, 15)	0.000859	0.068373	0.933912	58.733723	28.283057
[15, 16)	0.001423	0.069796	0.932584	57.783767	28.110830
[16, 17)	0.001423	0.071219	0.931258	56.865340	27.939739
[17, 18)	0.001423	0.072642	0.929934	55.945605	27.763724
[18, 19)	0.001423	0.074065	0.928611	55.024560	27.582493
[19, 20)	0.001423	0.075488	0.927291	54.102204	27.395704
[20, 21)	0.001851	0.077339	0.925576	53.178534	27.203509
[21, 22)	0.001851	0.079190	0.923864	52.276132	27.010599
[22, 23)	0.001851	0.081041	0.922156	51.372058	26.812161
[23, 24)	0.001851	0.082892	0.920451	50.466310	26.608025
[24, 25)	0.001851	0.084743	0.918748	49.558883	26.398016
[25, 26)	0.001859	0.086602	0.917042	48.649775	26.181953
[26, 27)	0.001859	0.088461	0.915339	47.739369	25.959743
[27, 28)	0.001859	0.090320	0.913639	46.827269	25.731103
[28, 29)	0.001859	0.092179	0.911942	45.913471	25.495831
[29, 30)	0.001859	0.094038	0.910248	44.997973	25.253722
[30, 31)	0.002022	0.096060	0.908410	44.080772	25.003902
[31, 32)	0.002022	0.098082	0.906575	43.168981	24.748766
[32, 33)	0.002022	0.100104	0.904743	42.255345	24.486184
[33, 34)	0.002022	0.102126	0.902916	41.339860	24.215924
[34, 35)	0.002022	0.104148	0.901092	40.422522	23.937742
[35, 36)	0.002420	0.106568	0.898914	39.503327	23.651389
[36, 37)	0.002420	0.108988	0.896741	38.597829	23.361665
[37, 38)	0.002420	0.111408	0.894574	37.690138	23.063437
[38, 39)	0.002420	0.113828	0.892411	36.780247	22.756430
[39, 40)	0.002420	0.116248	0.890254	35.868151	22.440363
[40, 41)	0.002986	0.119234	0.887600	34.953846	22.114942
[41, 42)	0.002986	0.122220	0.884954	34.056879	21.786612
[42, 43)	0.002986	0.125206	0.882315	33.157229	21.448045

Table C.2 (Continued)

[43, 44)	0.002986	0.128192	0.879684	32.254889	21.099412
[44, 45)	0.002986	0.131178	0.877062	31.349851	20.740377
[45, 46)	0.004379	0.135557	0.873229	30.442107	20.370597
[46, 47)	0.004379	0.139936	0.869414	29.573511	20.007475
[47, 48)	0.004379	0.144315	0.865615	28.701103	19.633467
[48, 49)	0.004379	0.148694	0.861833	27.824866	19.248188
[49, 50)	0.004379	0.153073	0.858067	26.944784	18.851239
[50, 51)	0.006497	0.159570	0.852510	26.060840	18.441359
[51, 52)	0.006497	0.166067	0.846989	25.227449	18.044173
[52, 53)	0.006497	0.172564	0.841504	24.388626	17.634787
[53, 54)	0.006497	0.179061	0.836055	23.544336	17.212727
[54, 55)	0.006497	0.185558	0.830641	22.694542	16.777501
[55, 56)	0.010937	0.196495	0.821605	21.839210	16.328592
[56, 57)	0.010937	0.207432	0.812669	21.073878	15.917188
[57, 58)	0.010937	0.218369	0.803829	20.300129	15.489973
[58, 59)	0.010937	0.229306	0.795085	19.517872	15.048848
[59, 60)	0.010937	0.240243	0.786437	18.727012	14.593161
[60, 61)	0.018857	0.259100	0.771746	17.927455	14.122227
[61, 62)	0.018857	0.277957	0.757329	17.259202	13.717924
[62, 63)	0.018857	0.296814	0.743182	16.578229	13.298268
[63, 64)	0.018857	0.315671	0.729299	15.884293	12.863667
[64, 65)	0.018857	0.334528	0.715676	15.177147	12.409777
[65, 66)	0.030552	0.365080	0.694141	14.456540	11.936782
[66, 67)	0.030552	0.395632	0.673254	13.889521	11.559724
[67, 68)	0.030552	0.426184	0.652996	13.304911	11.161633
[68, 69)	0.030552	0.456736	0.633348	12.702164	10.745108
[69, 70)	0.030552	0.487288	0.614290	12.080718	10.309238
[70, 71)	0.050173	0.537461	0.584230	11.439992	9.848062
[71, 72)	0.050173	0.587634	0.555640	11.002887	9.532289
[72, 73)	0.050173	0.637807	0.528450	10.543292	9.190369
[73, 74)	0.050173	0.687980	0.502590	10.060049	8.830791
[74, 75)	0.050173	0.738153	0.477996	9.551942	8.442710
[75, 76)	0.074639	0.812792	0.443618	9.017692	8.033252
[76, 77)	0.074639	0.887431	0.411712	8.677771	7.765322
[77, 78)	0.074639	0.962070	0.382101	8.311509	7.476191
[78, 79)	0.074639	1.036709	0.354620	7.916862	7.162684
[79, 80)	0.074639	1.111348	0.329115	7.491633	6.816934
[80, 81)	0.111104	1.222452	0.294507	7.033450	6.444389
[81, 82)	0.111104	1.333556	0.263538	6.801203	6.250702
[82, 83)	0.111104	1.444660	0.235826	6.541665	6.033343
[83, 84)	0.111104	1.555764	0.211028	6.251627	5.790441
[84, 85)	0.111104	1.666868	0.188838	5.927508	5.517175
[85, 86)	0.149720	1.816588	0.162580	5.565300	5.204731
[86, 87)	0.149720	1.966308	0.139973	5.383392	5.047815
[87, 88)	0.149720	2.116028	0.120509	5.172103	4.865556
[88, 89)	0.149720	2.265748	0.103752	4.926690	4.651963
[89, 90)	0.149720	2.415468	0.089326	4.641640	4.398700
[90, 91)	0.200680	2.616148	0.073084	4.310552	4.104533
[91, 92)	0.200680	2.816828	0.059795	4.157384	3.968445

Table C.2 (Continued)

[92, 93)	0.200680	3.017508	0.048923	3.970178	3.801320
[93, 94)	0.200680	3.218188	0.040028	3.741368	3.591927
[94, 95)	0.200680	3.418868	0.032749	3.461709	3.335999
[95, 96)	0.263290	3.682158	0.025169	3.119901	3.023196
[96, 97)	0.263290	3.945448	0.019343	2.909025	2.827717
[97, 98)	0.263290	4.208738	0.014865	2.634633	2.569076
[98, 99)	0.263290	4.472028	0.011424	2.277593	2.232532
[99, 100)	0.263290	4.735318	0.008780	1.813012	1.789332
[100, 101)	0.344611	5.079929	0.006220	1.208496	1.202423

Notes: All-cause mortality rates are from published life tables for females in India (WHO 2013). Hazard rate h(t) denotes the rate that applies within the interval [t, t+1). Formulas for the remaining columns are as follows:

- Cumulative hazard at age t+1 years: $H(t+1) = h(0) + h(1) + \ldots + h(t)$.
- Probability of surviving to age t+1 years: $S(t+1)=e^{H(t+1)}$.
- Remaining life expectancy from age t years: $LE(t) = [S(t+1) + \ldots + S(101)]/S(t) + 0.5$. (The addition of 0.5 reflects the assumption that deaths occur at the midpoint of each interval.)
- Discounted life expectancy from age t, disc_LE(t), is computed by discounting future life years by 3% per year. For example, $disc\_LE(23) = 1 + 1/1.03 + 1/1.03^2 + \ldots + 1/1.03^{49} + 0.46631/1.03^{50}$.

C.3: Details on P4P program cost components

Table C.3: Description of program activities and expenditures

Program task	Description of inputs
SET-UP COSTS	
Production of survey instruments for P4P program	Time of study investigators spent developing questionnaire forms for surveys of providers/personnel and households, and associated travel and materials. Activities included drafting the instruments, piloting, focus groups, panel discussions, a validation study, and finalizing the instruments.
Identification of rural geographic areas to target	Cartography work performed by GIS Lab using 2001 census data from the Government of Karnataka to generate maps of hoblis with no large public health provider.
Identification of potentially eligible providers within targeted regions	Government-commissioned fieldwork to collect initial list of 319 potentially eligible providers. Involved interviews with local informants to identify all formal medical providers offering obstetric services, and collection of basic demographic and infrastructure data. (Cost estimate assumes that the unit cost of collecting basic data on each of the 319 providers was roughly equal to one-half of the unit cost per in-person provider visit.)
Screening of ineligible providers	Interviews with providers who were determined to be ineligible for P4P program participation. (Cost estimate is based on the numbers of in-person visits conducted among providers who were excluded, times the unit cost per in-person visit.)
RECURRENT COSTS	
Provider payouts	
Participation rewards	Three installments of Rs. 2,500 paid as compensation for participating in the program.
Record-keeping reward	A nominal sum of Rs. 1,000 to each provider for the additional record-keeping effort required to transmit patient lists to the study team.
Input P4P performance reward	Average reward payout to providers under input-based P4P in the experiment.
Outcome P4P performance reward	Average reward payout to providers under outcome-based P4P in the experiment.
Provider visits & household data collection	
Training of field associates	Wages of field managers and trainees, venue costs, and training materials. Field associates are responsible for conducting meetings with providers and helping to monitor the large-scale household data collection effort. (Costs of training household interviewers are captured in household data collection costs below.)
In-person provider visits	Includes the expenses associated with: field staff wages, transportation, and meals/incidentals; office supplies and printed materials for meetings; dataset compilation; administrative support; organizational overhead; and oversight of the field work by managerial staff.
Household data collection	Covers the cost of interviewing approximately 19 households per provider during the year to assess performance. The unit cost of household data collection includes similar expense categories as the in-person provider visits above. The costs of recruiting and training household interviewer staff are also built into this unit cost, as these training/recruitment costs were not recorded separately in project invoices.
External supervision and field monitoring support	Wages, housing allowance, travel expenses, and other costs associated with employing a full-time project manager hired by the impact evaluation team to supervise the field work by partners at Sambodhi and ensure adherence to data collection protocols.

C.4: Potential impact of provider incentives on patient-reported hospital costs

One potential unintended consequence of the P4P initiatives is an increase in patients' out-of-pocket costs: Obstetric care providers in the region (who generally bill according to fee-for-service) could conceivably pass along the cost of additional incentivized services to their patients. I investigate this possibility through regression analyses that estimate the impact of provider incentives on patient-reported costs of delivery. If input or outcome P4P show evidence of increasing patient costs relative to control contracts, a conservative approach would be to include this cost increase when measuring the cost consequences of P4P from a societal perspective, rather than relying on performance rewards alone to approximate the cost of providers' quality improvement efforts.

On average, households reported paying a total hospital bill of Rs. 13,288 (~$218) for delivery at participating providers' facilities. Table C.4 summarizes results from linear regression models with varying levels of covariate adjustment. As shown, hospital costs are numerically (though non-significantly) lower under input and outcome P4P compared to control contracts in the unadjusted model (column [i]) and when adjusting for exogenous provider- and patient-level characteristics (column [ii]). Similar results are obtained when also adjusting for the occurrence of maternal complications, neonatal mortality, and/or caesarean mode of delivery (columns [iii]-[iv]), suggesting that the numerically lower delivery costs under P4P are not explained by differences in birth outcomes between the groups.

These results provide some support for the base-case model assumption that performance rewards adequately compensated providers for their quality improvement efforts. However, as discussed in the main text, it is still possible that providers themselves absorbed the cost of

uncompensated effort (e.g., due to altruism or because providers tended to expect higher rewards than they received).

Table C.4: Impact of provider incentives on patient-reported hospital costs: Multivariate linear regression results

	Cost difference in INR (95% CI), by model specification			
	(i) Unadjusted	(ii) Adjusted for baseline covariates only	(iii) Adjusted for baseline covariates and complications	(iv) Adjusted for baseline covariates, complications, and delivery mode
Treatment group				
Input P4P	-1175 (-4548, 2198)	-1431 (-4334, 1472)	-1466 (-4465, 1534)	-1629 (-4308, 1050)
Outcome P4P	-134 (-3247, 2979)	-984 (-3389, 1421)	-1052 (-3407, 1303)	-993 (-3012, 1026)
Control	(ref)	(ref)	(ref)	(ref)
Complication(s)	N	N		
Postpartum hemorrhage			312 (-1534, 2158)	1001 (-813, 2815)
Pre-eclampsia/eclampsia			151 (-1454, 1756)	378 (-1065, 1821)
Sepsis			1778 (-1431, 4987)	2661 (-515, 5838)
Neonatal mortality			3706 (-5157, 12569)	8048 (-1167, 17264)
Delivery mode	N	N	N	
Caesarean section				14285 (11897, 16674)
Vaginal delivery				(ref)
Provider-level controls	N	Y	Y	Y
Patient-level controls	N	Y	Y	Y
N	2,536	2,536	2,536	2,536

Notes: Results are from linear regression models with clustered standard errors at the provider level. Dependent variable is self-reported total bill for delivery at the participating provider's facility. Cost values are missing for 72 (2.8%) of the 2,608 surveyed households.

C.5: Estimates of direct medical costs per maternal complication case

Table C.5: Base-case estimates of cost per case derived from the OneHealth costing tool and WHO CHOICE databases

Complication / Cost category	2014 Rs	2014 USD
Postpartum hemorrhage		
Drugs & supplies	702.33	11.51
Personnel	1,268.26	20.78
Facility	2,669.26	43.74
Total	**4,639.84**	**76.03**
Pre-eclampsia/eclampsia		
Drugs & supplies	296.08	4.85
Personnel	2,555.32	41.87
Facility	3,736.96	61.23
Total	**6,588.35**	**107.95**
Sepsis		
Drugs & supplies	4,714.22	77.24
Personnel	290.50	4.76
Facility	2,135.40	34.99
Total	**7,140.13**	**116.99**

Notes: Drugs & supplies cost is directly estimated in the OneHealth costing tool using an ingredients approach. I derive the personnel and facility costs using resource input estimates from the OneHealth tool (i.e., average personnel minutes and length of hospital stay required per complication case), and unit costs for these resource inputs from WHO CHOICE. Unit cost estimates are provided in Table C.6 below.

Table C.6: Unit costs of medical personnel time and hospital days from WHO CHOICE

Cost type	Unit cost estimate	
	2014 Rs	2014 USD
Annual wages, by type of medical personnel		
OB/GYN	808,608	13,249
Anesthetist	808,608	13,249
Midwives	174,992	2,867
Assistant nurses and midwives	122,760	2,011
Laboratory technicians/assistants	174,992	2,867
Cost per bed day at a primary-level hospital	534	8.75

Notes: Unit costs estimates are extracted from the WHO CHOICE databases and adjusted to 2014 currency units using GDP deflators for India. In order to value the OneHealth model's estimates of personnel minutes required per complication case, the annual personnel wages listed above are divided out with the assumption of a 40-hour work week and 49 work weeks per year. The unit cost per hospital bed day estimates the "hotel" portion of hospital costs only, thus avoiding overlap with the personnel and drugs/supplies cost components.

Table C.7: Direct costs per complication case: Alternative literature estimates

Complication / Reference	Setting	2014 Rs[a]	2014 USD
Postpartum hemorrhage			
Levin 2003[b]	Ghana - Public hospital	8,896.06	145.77
	Ghana - Mission hospital	3,596.14	58.92
	Malawi - Public hospital	7,802.00	127.84
	Malawi - Missing hospital	6,425.57	105.29
	Uganda - Public hospital	4,846.22	79.41
	Uganda - Mission hospital	10,991.34	180.10
Borghi 2003[c]	Benin - Teaching hospital	12,075.27	197.86
	Benin - Non-teaching hospital	8,666.93	142.01
	Ghana - Teaching hospital	8,861.69	145.20
	Ghana - Non-teaching hospital	5,453.35	89.36
Weissman 1999[d]	Uganda - Hospital	3,224.79	52.84
Pre-eclampsia/eclampsia			
Levin 2003[b]	Malawi - Public hospital	10,201.66	167.16
	Malawi - Missing hospital	5,040.53	82.59
	Uganda - Public hospital	7,884.32	129.19
	Uganda - Mission hospital	15,282.39	250.41
Borghi 2003[c]	Benin - Teaching hospital	14,120.27	231.37
	Benin - Non-teaching hospital	10,127.64	165.95
	Ghana - Teaching hospital	11,004.07	180.31
	Ghana - Non-teaching hospital	4,479.53	73.40
Weissman 1999[d]	Uganda - Hospital	5,127.45	84.02
Sepsis			
Borghi 2003[c]	Benin - Teaching hospital	20,839.57	341.47
	Benin - Non-teaching hospital	8,569.54	140.42
	Ghana - Teaching hospital	7,790.50	127.65
Weissman 1999[d]	Uganda - Hospital	788.91	12.93

[a] Costs from each publication are first translated from USD into Rs. units using exchange rates for the original reporting year, and then inflation-adjusted to 2014 Rs. using GDP deflators.
[b] Levin (2003) interviewed personnel at six hospitals in Ghana, Malawi, or Uganda to estimate direct medical costs per complication case.
[c] Borghi (2003) conducted a survey of women who gave birth at five referral hospitals in Benin or Ghana to estimate direct costs (medical + non-medical) incurred by patients and their relatives. Note that the costs originally reported in the paper correspond to total costs associated with complicated deliveries, rather than the marginal cost of complication management (i.e., above and beyond normal delivery expenses). Therefore, to estimate the marginal cost associated with a given complication, I subtract the reported average cost of a normal delivery from the average cost of a delivery with that complication.
[d] Weissman (1999) conducted site visits and interviews with staff at two hospitals in Uganda to estimate direct medical costs associated with complication management.